

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:17:14 ; Search time 50.5 Seconds
(without alignments)
1818.375 Million cell updates/sec

Title: US-10-720-192-2
Perfect score: 1726
Sequence: 1 MHQVDNLTTRKGRLLAALAI.....PAGEVAPTTPTPTQTLPA 325

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1726	100.0	325	2	AAW03565	Aaw03565 Mycobacte
2	1726	100.0	325	5	AAW50732	Aam50732 Mycobacte
3	1726	100.0	332	2	AAW32350	Aaw32350 Mycobacte
4	1726	100.0	332	2	AAW32418	Aaw32418 Mycobacte
5	1726	100.0	332	2	AAW43322	Aaw43322 Mycobacte
6	1726	100.0	332	2	AAW81683	Aaw81683 M. tuberc
7	1726	100.0	332	2	AAW38945	Aay38945 M. tuberc
8	1726	100.0	332	2	AAW39083	Aay39083 M. tuberc
9	1726	100.0	332	4	AAU01895	Aau01895 M. tuberc
10	1726	100.0	332	5	AAE29720	Aae29720 Mycobacte
11	1726	100.0	332	5	AAE17584	Aae17584 Mycobacte
12	1720	99.7	325	3	AAW37281	Aay37281 Fibronect
13	1542	89.3	286	2	AAW03566	Aaw03566 Mycobacte
14	1499	86.8	652	2	AAW39082	Aay39082 M. tubercu
15	1499	86.8	652	2	AAW39225	Aay39225 M. tuberc
16	1499	86.8	802	2	AAW64379	Aaw64379 Mycobacte
17	1499	86.8	802	2	AAW81746	Aaw81746 M. tuberc
18	1499	86.8	802	2	AAW32063	Aay32063 Mycobacte
19	1499	86.8	802	2	AAW39033	Aay39033 M. tuberc
20	1499	86.8	802	2	AAW39081	Aay39081 M. tubercu
21	1499	86.8	802	2	AAW39176	Aay39176 M. tuberc
22	1499	86.8	802	2	AAW39224	Aay39224 M. tuberc
23	1499	86.8	802	2	AAU74592	Aau74592 Antigenic
24	980	56.8	287	7	ADB74347	Adb74347 Mycobacte
25	237.5	13.8	268	4	ABU52942	Abu52942 Human mam

26	235	13.6	256	4	ABU52945	Abu52945 Human mam
27	233.5	13.5	262	4	ABU52944	Abu52944 Human mam
28	233	13.5	270	4	ABU52939	Abu52939 Human mam
29	233	13.5	8991	6	ABU08487	Abu08487 S. pneumo
30	228	13.2	544	3	AAG15453	Aag15453 Arabidops
31	227.5	13.2	652	7	ADB65210	Adb65210 Human pro
32	227	13.2	598	6	ABU37030	Abu37030 Protein e
33	225.5	13.1	446	4	ABW70063	Abw70063 Drosophil
34	223.5	12.9	652	5	ABB92424	Abb92424 Herbicida
35	222.5	12.9	267	4	ABU52938	Abu52938 Human mam
36	222	12.9	763	2	AAW31852	Aaw31852 Mycobacte
37	221.5	12.8	260	4	ABU52932	Abu52932 Human mam
38	221.5	12.8	267	4	ABU52931	Abu52931 Human mam
39	218.5	12.7	253	4	ABU52937	Abu52937 Human mam
40	217.5	12.6	817	6	ABR53281	AbR53281 Protein s
41	217	12.6	261	4	ABU52940	Abu52940 Human mam
42	217	12.6	277	4	ABU52936	Abu52936 Human mam
43	216.5	12.5	254	4	ABU52933	Abu52933 Human mam
44	214	12.4	406	4	ABG27250	Abg27250 Novel hum
45	212.5	12.3	258	4	ABU52941	Abu52941 Human mam

ALIGNMENTS

RESULT 1
ID AAW03565 standard; protein; 325 AA.
XX
AC AAW03565;
XX
DT 22-APR-1997 (first entry)
XX
DE Mycobacterium tuberculosis antigenic determinant protein.
XX
KW Antigenic determinant; Mycobacterium tuberculosis; ion exchange; human;
KW chromatography; gel filtration; reverse phase column chromatography;
KW immunogenic; serum; guinea pig; expression vector; cosmid; antibody;
KW Mycobacterium smegmatis; Mycobacterium bovis BCG; microorganism; vaccine;
KW hybrid; epitope; disease; diphtheria; cholera; toxin.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Peptide 1..39
FT Protein /note= "signal peptide"
FT /note= "mature protein"
XX
XX WO9623885-A1.
XX
PD 08-AUG-1996.
XX
PF 31-JAN-1996; 96WO-FR000166.
XX
PR 01-FEB-1995; 95US-00382184.
XX
XX (INSP) INST PASTEUR.
XX
PI Laqueyzerie A, Marchal G, Pescher P, Romain F;
XX
XX WPI; 1996-371433/37.
XX
DR N-PSDB; AAT39357.
XX
XX Mycobacterium tuberculosis antigen and hybrid proteins comprising it -
XX useful in vaccines against tuberculosis and in immunoassays.
XX
XX Claim 2; Page 49-50; 74pp; French.
XX
XX This is the amino acid sequence of an antigenic determinant protein from
XX Mycobacterium tuberculosis. The mature protein has calculated mol. wt. of
XX 28779 Da but has an experimental mol. wt of 45-47 kD as determined by SDS
XX -PAGE. The difference is thought to be due to the high frequency of Pro

Db	248	GVIGSPANAPDAGPQRFVFWVLTGTTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP	307
QY	301	APAPAPAGEVAPTPTTPTPQRTLPA	325
Db	308	APAPAPAGEVAPTPTTPTPQRTLPA	332
RESULT 5			
ID	AAW64322		
XX	AAW64322	standard; protein; 332 AA.	
XX	AAW64322;		
XX	AC		
XX	AC		
DT	17-OCT-2003	(revised)	
DT	09-NOV-1998	(first entry)	
XX	XX		
DE	XX	Mycobacterium tuberculosis antigen DPEP.	
XX	XX	Tuberculosis; infection; diagnosis; antigen; DPEP.	
XX	XX	Mycobacterium tuberculosis; strain H37Rv.	
OS	XX		
PN	XX	WC9816645-A2.	
PD	XX		
PD	XX	23-APR-1998.	
XX	XX		
XX	XX	07-OCT-1997; 97WO-US018214.	
XX	XX		
PR	XX	11-OCT-1996; 96US-00729622.	
PR	XX	13-MAR-1997; 97US-00818111.	
XX	XX		
PA	XX	(CORI-) CORIXA CORP.	
PI	XX		
PI	XX	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;	
XX	XX	Vedvick TS, Twardzik DR, Lodes MJ;	
DR	XX		
DR	XX	WPI; 1998-251292/22.	
DR	XX	N-PSDB; AAW44390.	
XX	XX		
PT	XX	New isolated Mycobacterium tuberculosis polypeptides and DNA - used to	
PT	XX	develop products for the detection of M. tuberculosis infection and	
PT	XX	diagnosis of tuberculosis.	
XX	XX		
PS	XX	Example 1; Page 93-94; 250pp; English.	
XX	XX		
CC	XX	This polypeptide comprises the Mycobacterium tuberculosis antigen DPEP. A	
CC	XX	DNA sequence (see AAW44390) coding for DPEP was isolated from a M.	
CC	XX	tuberculosis genomic library using a probe based on an isolated N-	
CC	XX	terminal peptide (see AAW64347). The invention relates to compositions	
CC	XX	and methods for diagnosing tuberculosis. It provides polypeptides (see	
CC	XX	AAW64291-W64379) comprising an antigenic portion of a soluble M.	
CC	XX	tuberculosis antigen, or an immunogenic portion of an M. tuberculosis	
CC	XX	antigen, as well as DNA sequences encoding such polypeptides, recombinant	
CC	XX	expression vectors and transformed or transfected host cells. Also	
CC	XX	claimed are methods and diagnostic kits for detecting M. tuberculosis	
CC	XX	infection in a patient using these polypeptides, antibodies or	
CC	XX	oligonucleotide probes and primers, for the diagnosis of tuberculosis.	
CC	XX	(Updated on 17-OCT-2003 to standardise OS field)	
XX	XX		
SQ	XX	Sequence 332 AA;	
Query Match		100.0%; Score 1726; DB 2; Length 332;	
Best Local Similarity		100.0%; Pred. No. 1.3e-104;	
Matches 325; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MEQVDNLTTRKGRLLAALIAAASASLTVVAVPATANADPEPAPVPTTAAASPPSTAAA	60
Db	8	MEQVDNLTTRKGRLLAALIAAASASLTVVAVPATANADPEPAPVPTTAAASPPSTAAA	67
QY	61	PPAPATPVAPPPPPAAANTTNAQPGDFAAAPPADFNAPPPPVVIAPNAPQPVRIIDNPVGGF	120
Db	68	PPAPATPVAPPPPPAAANTTNAQPGDFAAAPPADFNAPPPPVVIAPNAPQPVRIIDNPVGGF	127

QY	121	SPALPAGWVESDAAHFYGSGALLSKTTGDPFFGQPPPVANDTRIIVLGRLDQKLYASAEA	187			
DB	128	SPALPAGWVESDAAHFYGSGALLSKTTGDPFFGQPPPVANDTRIIVLGRLDQKLYASAEA	187			
QY	181	TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPGQIWT	240			
DB	188	TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPGQIWT	247			
QY	241	GVIGSPAANADGAPQGRFVWVLGTANNPVDKGAALAESTRPLVAPPPAPAPAPAP	300			
DB	248	GVIGSPAANADGAPQGRFVWVLGTANNPVDKGAALAESTRPLVAPPPAPAPAPAP	307			
QY	301	APAPAPAGEVAPTTPTTPTQRTLEA	325			
DB	308	APAPAPAGEVAPTTPTTPTQRTLEA	332			
RESULT 6						
AAW81683						
ID	AAW81683 standard; protein; 332 AA.					
XX	AAW81683;					
DT	27-JAN-1999 (first entry)					
XX	M. tuberculosis immunogenic polypeptide DPEP.					
DE	Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;					
XX	vaccine; pharmaceutical; infection; diagnosis.					
KW	Mycobacterium tuberculosis.					
XX	WO9816646-A2.					
PN	23-APR-1998.					
PD	07-OCT-1997; 97WO-US018293.					
XX	11-OCT-1996; 96US-00730510.					
PR	13-MAR-1997; 97US-00818112.					
XX	(CORI-) CORIXA CORP.					
PA	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;					
PI	Weddick TS, Twardzik DR, Lodes MJ;					
PI	WPI: 1998-261042/23.					
XX	N-PSDB; AAV64498.					
DR	Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to					
XX	develop products for the detection of M. tuberculosis infection and for					
PT	diagnosis, treatment and prevention of tuberculosis.					
PT	Example 1; Page 91-92; 230pp; English.					
PS	This sequence represents an immunogenic portion of a soluble					
XX	Mycobacterium tuberculosis (MT) antigen which can be used in a method for					
CC	inducing protective immunity against tuberculosis (TB). This sequence can					
CC	be formulated into vaccines and/or pharmaceutical compositions for					
CC	immunising against M. tuberculosis infection or may be used for the					
CC	diagnosis of tuberculosis					
XX	Sequence 332 AA;					
SQ	Query Match 100.0%; Score 1726; DB 2; Length 332;					
Best Local Similarity 100.0%; Pred. No. 1.3e-104;						
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0						
QY	1	MHQVDPNLTRKGLAALATAAMASALVTVAVPATANADPEPAPVPTTAASPPSTAAA	60			
DB	8	MHQVDPNLTRKGLAALATAAMASALVTVAVPATANADPEPAPVPTTAASPPSTAAA	67			
QY	61	PPAPATPVAPPPPPAAANTPNAQGDGPNNAAPFPADPNAPPPFVTIAPNAPQFVIDNPVGGF	120			

QY	121	SPALPAGWVESDAAHFYDYGSA	LLSKTTGDPFFGQPPPVANDTRI	VLGRLDQKLYASAEA	187					
Db	128	SPALPAGWVESDAAHFYDYGSA	LLSKTTGDPFFGQPPPVANDTRI	VLGRLDQKLYASAEA	187					
QY	181	TDSKAAARLGSDMGGEFYMPY	PGTRINQETVSLDANGVSGSAS	YVEVKFSDPSKPGQIWT	240					
Db	188	TDSKAAARLGSDMGGEFYMPY	PGTRINQETVSLDANGVSGSAS	YVEVKFSDPSKPGQIWT	247					
QY	241	GVIGSPAANADGAPQORFWV	VLGTANNPVDGAAKALAESIR	PLVAPPPAPAPAPAP	300					
Db	248	GVIGSPAANADGAPQORFWV	VLGTANNPVDGAAKALAESIR	PLVAPPPAPAPAPAP	307					
QY	301	APAPAPAGEVAPTTPTT	PQRTPLA	325						
Db	308	APAPAPAGEVAPTTPTT	PQRTPLA	332						
RESULT 6										
AAW81683										
ID	AAW81683 standard; protein; 332 AA.									
XX	AAW81683;									
DT	27-JAN-1999 (first entry)									
XX	M. tuberculosis immunogenic polypeptide DPEP.									
DE	Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;									
XX	vaccine; pharmaceutical; infection; diagnosis.									
KW	Mycobacterium tuberculosis.									
XX	W09816646-A2.									
PN	23-APR-1998.									
PD	07-OCT-1997; 97WO-US018293.									
XX	11-OCT-1996; 96US-00730510.									
PR	13-MAR-1997; 97US-00818112.									
XX	(CORI-) CORIXA CORP.									
PA	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;									
PI	Vedick TS, Twardzik DR, Lodes MJ;									
PI	WPI: 1998-261042/23.									
XX	N-PSDB; AAV64498.									
DR	Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to									
XX	develop products for the detection of M. tuberculosis infection and for									
PT	diagnosis, treatment and prevention of tuberculosis.									
PT	Example 1; Page 91-92; 230pp; English.									
PS	This sequence represents an immunogenic portion of a soluble									
XX	Mycobacterium tuberculosis (MT) antigen which can be used in a method for									
CC	inducing protective immunity against tuberculosis (TB). This sequence can									
CC	be formulated into vaccines and/or pharmaceutical compositions for									
CC	immunising against M. tuberculosis infection or may be used for the									
CC	diagnosis of tuberculosis									
XX	Sequence 332 AA;									
SQ	Query Match 100.0%; Score 1726; DB 2; Length 332;									
Best Local Similarity 100.0%; Pred. NO. 1.3e-104;										
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0										
QY	1	MHQVDPNLTRKGLAALATAAMAS	SLVTVAVPATANADPEPAPVPTT	AAASPPSTAAA	60					
Db	8	MHQVDPNLTRKGLAALATAAMAS	SLVTVAVPATANADPEPAPVPTT	AAASPPSTAAA	67					
QY	61	PPAPATPVAPPPPPAAANTPNA	QGDGDNAAFPFPADPNAPFP	FPVTIAPNAPQFVIDNPVG	120					

Db	68	PPAPATPVAPPFPFAAANTPNAQPGDPNAAFPADPNAPPVFIAPNAPQFVRIDNPVGGF	127			
Qy	121	SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRI VLGRLDOKLYASAEA	180			
Db	128	SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRI VLGRLDOKLYASAEA	187			
Qy	181	TTSKAAARLGSDMGFBFYMPYPCGTRINQETVSLDANGVSGSASYEVKFSDFSPKNGQIWT	240			
Db	186	TTSKAAARLGSDMGFBFYMPYPCGTRINQETVSLDANGVSGSASYEVKFSDFSPKNGQIWT	247			
Qy	241	GVIGSPAANAPDAGPQQPQRFVVMGLGTANNPVDKGAAXALAESIRPLVAPPAPAPAPAE	300			
Db	248	GVIGSPAANAPDAGPQQPQRFVVMGLGTANNPVDKGAAXALAESIRPLVAPPAPAPAPAE	307			
Qy	301	APAPAPAGEVAFTPTTPTPQRTLPA	325			
Db	308	APAPAPAGEVAFTPTTPTPQRTLPA	332			
RESULT 7						
AA38945						
ID	AA38945 standard; protein; 332 AA.					
XX	AA38945;					
XX	XX					
XX	XX					
DT	05-NOV-1999 (first entry)					
XX	M. tuberculosis recombinant antigen protein DPEP.					
DE	XX					
XX	XX					
KW	Antigen; diagnosis; detection; infection; antibody; immunisation;					
KW	vaccine; immunity.					
XX	XX					
OS	Mycobacterium tuberculosis.					
XX	XX					
PN	W09942118-A2.					
XX	XX					
PD	26-AUG-1999.					
XX	XX					
PF	17-FEB-1999; 99WO-US003265.					
XX	XX					
PR	18-FEB-1998; 98US-00024753.					
PR	05-MAY-1998; 98US-00072596.					
XX	XX					
PA	(CORI-) CORIXA CORP.					
XX	XX					
PI	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;					
PI	Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;					
XX	XX					
DR	WPI; 1999-527416/44.					
DR	N-PSDB; AA319088.					
XX	XX					
PT	New polypeptide comprising antigenic portions of M. tuberculosis.					
XX	XX					
PS	Example 1; Page 131-132; 323pp; English.					
XX	XX					
CC	This invention describes novel recombinant antigens and their encoding					
CC	nucleic acids derived from Mycobacterium tuberculosis. The novel					
CC	polypeptides are useful for detecting M. tuberculosis infection in a					
CC	biological sample by detecting antibodies which bind with the					
CC	polypeptides, and are useful as vaccines for immunizing against M.					
CC	tuberculosis infection. The new detection methods are needed as current					
XX	vaccination strategies do not provide 100% immunity					
XX	XX					
SQ	Sequence 332 AA;					
Query Match						
Best Local Similarity 100.0%; Score 1726; DB 2; Length 332;						
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0						
Qy	1	MHVDPNLTFRKGRILAALAIAMASASLTVAVPATANADPEPAPVPTTAASPPSTAAA	60			
Db	8	MHVDPNLTFRKGRILAALAIAMASASLTVAVPATANADPEPAPVPTTAASPPSTAAA	67			

Query Match 100.0%; Score 1726; DB 2; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e-104;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
 |||||
 DB 8 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 67
 |||||

QY 61 PPAPATPVAPPPPPAAANTPNAPQGDPAAPPPADPNAPPVPIAPNAPQPVRIIDNPVGGF 120
 |||||
 DB 68 PPAPATPVAPPPPPAAANTPNAPQGDPAAPPPADPNAPPVPIAPNAPQPVRIIDNPVGGF 127
 |||||

QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPVANDTRIVLGRDQKLYASAEA 180
 |||||
 DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPVANDTRIVLGRDQKLYASAEA 187
 |||||

QY 181 TDSKAAARLGSDMGDFYMPYPCPTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 240
 |||||
 DB 188 TDSKAAARLGSDMGDFYMPYPCPTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 247
 |||||

QY 241 GYIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPAPAPAPAPAP 300
 |||||
 DB 248 GYIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPAPAPAPAPAP 307
 |||||

QY 301 APAPAPAGEVAPTPTTPTPQRTPLA 325
 |||||
 DB 308 APAPAPAGEVAPTPTTPTPQRTPLA 332
 |||||

RESULT 9
 AAU01895
 ID AAU01895 standard; protein; 332 AA.
 AC AAU01895;
 DT 29-AUG-2001 (first entry)
 DE M. tuberculosis DPEP antigen.
 DE DPEP; antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease.
 XX Mycobacterium tuberculosis.
 OS WO200124820-A1.
 PN 12-APR-2001.
 XX 10-OCT-2000; 2000WO-US028095.
 XX 07-OCT-1999; 99US-0158338P.
 PR 07-OCT-1999; 99US-0158425P.
 XX (CORI-) CORIXA CORP.
 PA Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 DR N-PSDB; AAS03786.
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens.
 XX Disclosure; Page 160; 168pp; English.
 XX The sequence represents Mycobacterium tuberculosis DPEP, an M.
 CC tuberculosis antigen. Compositions comprising at least 2 heterologous
 CC antigens, as a fusion protein, and vectors expressing the fusion proteins
 CC are used as vaccines to prophylactically immunise mammals (especially
 CC humans) against infection by Mycobacteria. The compositions contain at
 CC least 2 heterologous antigens that increase the serological sensitivity
 CC of individuals infected with tuberculosis, a disease frequently affecting

CC patients with acquired immunodeficiency disease, AIDS
 XX
 SQ Sequence 332 AA;
 Query Match 100.0%; Score 1726; DB 4; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e-104;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
 |||||
 DB 8 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 67
 |||||

QY 61 PPAPATPVAPPPPPAAANTPNAPQGDPAAPPPADPNAPPVPIAPNAPQPVRIIDNPVGGF 120
 |||||
 DB 68 PPAPATPVAPPPPPAAANTPNAPQGDPAAPPPADPNAPPVPIAPNAPQPVRIIDNPVGGF 127
 |||||

QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPVANDTRIVLGRDQKLYASAEA 180
 |||||
 DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPVANDTRIVLGRDQKLYASAEA 187
 |||||

QY 181 TDSKAAARLGSDMGDFYMPYPCPTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 240
 |||||
 DB 188 TDSKAAARLGSDMGDFYMPYPCPTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 247
 |||||

QY 241 GYIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPAPAPAPAPAP 300
 |||||
 DB 248 GYIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPAPAPAPAPAP 307
 |||||

QY 301 APAPAPAGEVAPTPTTPTPQRTPLA 325
 |||||
 DB 308 APAPAPAGEVAPTPTTPTPQRTPLA 332
 |||||

RESULT 10
 AAU29720
 ID AAU29720 standard; protein; 332 AA.
 AC AAU29720;
 DT 27-JAN-2003 (first entry)
 DE Mycobacterium tuberculosis DPEP antigenic protein.
 DE Vaccine; immunity; diagnostic agent; gene therapy; DPEP antigen.
 KW Mycobacterium tuberculosis.
 OS WO200272792-A2.
 PN 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 XX 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 PA Skeiky Y, Brannon M, Guderian J;
 XX WPI; 2002-759844/82.
 DR N-PSDB; AAD47097.
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 CC tuberculosis.
 XX Disclosure; Page 111; 155pp; English.
 PS The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a

CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is M.
CC tuberculosis DPEP antigenic protein
XX
SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 5; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.3e-104;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDNLTTRKGRGLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
DB 8 MHQVDNLTTRKGRGLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
QY 61 PPAPATPVAPPPAAANTENAPQGDENAPADPNAPPPVPIAPNAPQVRIIDNPVGGF 120
DB 68 PPAPATPVAPPPAAANTENAPQGDENAPADPNAPPPVPIAPNAPQVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQGPVANDTRIVLGRDLQKLYASAEA 180
DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQGPVANDTRIVLGRDLQKLYASAEA 187
QY 181 TDSKAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYVEVKFSDPKNGQIWT 240
DB 188 TDSKAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYVEVKFSDPKNGQIWT 247
QY 241 GVIGSPAANADAGPPQRFVWVLTANNPVYDKGAALAEIRPLVAPPAPAPAPAP 300
DB 248 GVIGSPAANADAGPPQRFVWVLTANNPVYDKGAALAEIRPLVAPPAPAPAPAP 307
QY 301 APAPAPAGEVAPTPTTPTPQRTLPA 325
DB 308 APAPAPAGEVAPTPTTPTPQRTLPA 332

RESULT 11
AAE17584
ID AAE17584 standard; protein; 332 AA.
XX AAE17584;
AC AAE17584;
DT 22-APR-2002 (first entry)
DE Mycobacterium species DPEP protein.
XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; DPEP protein.
XX Mycobacterium sp.
OS WO200198460-A2.
PN 27-DEC-2001.
PD 20-JUN-2001; 2001WO-US019959.
PF 20-JUN-2000; 2000US-00597796.
PR 01-FEB-2001; 2001US-0265737P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Reed S, Alderson M;
PI WPI; 2002-147798/19.
DR N-PSDB; AAD28355.
XX Composition comprising MTB39 antigen and MTB32A antigen from

PT Mycobacterium species, useful for eliciting immune response in a subject.
XX Claim 9; Page 127; 136pp; English.

CC The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC DPEP protein
XX
SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 5; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.3e-104;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDNLTTRKGRGLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
DB 8 MHQVDNLTTRKGRGLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
QY 61 PPAPATPVAPPPAAANTENAPQGDENAPADPNAPPPVPIAPNAPQVRIIDNPVGGF 120
DB 68 PPAPATPVAPPPAAANTENAPQGDENAPADPNAPPPVPIAPNAPQVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQGPVANDTRIVLGRDLQKLYASAEA 180
DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQGPVANDTRIVLGRDLQKLYASAEA 187
QY 181 TDSKAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYVEVKFSDPKNGQIWT 240
DB 188 TDSKAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYVEVKFSDPKNGQIWT 247
QY 241 GVIGSPAANADAGPPQRFVWVLTANNPVYDKGAALAEIRPLVAPPAPAPAPAP 300
DB 248 GVIGSPAANADAGPPQRFVWVLTANNPVYDKGAALAEIRPLVAPPAPAPAPAP 307
QY 301 APAPAPAGEVAPTPTTPTPQRTLPA 325
DB 308 APAPAPAGEVAPTPTTPTPQRTLPA 332

RESULT 12
AAAY97281
ID AAAY97281 standard; protein; 325 AA.
XX AAAY97281;
AC AAAY97281;
DT 03-JAN-2001 (first entry)
XX Fibronectin attachment protein FAP-B.
DE Inflammation; inflammatory response; irritant; pathogen; treatment;
KW T helper cell; lymphocyte; cell mediated immunity; skin allergy; hives;
KW allergic rhinitis; conjunctivitis; hay fever; allergic gastroenteritis;
KW asthma; bronchopulmonary aspergillosis; pollutant;
KW respiratory tract infection.

Db 181 SASYYEVKFSKNGQIWTGVIGSPAANAPDAGPPQRFVWMLGTANNPVDKGAAL 240
Qy 280 AESTIRPVAPPPAPAPAPAPAPAPAGEVAPTPTTTPQRTLP 325
Db 241 AESTIRPLVAPPPAPAPAPAPAPAGEVAPTPTTTPQRTLP 286

RESULT 14
AA39082
ID AA39082 standard; protein; 652 AA.
AC AA39082;
XX
XX
DT 05-NOV-1999 (first entry)
DE
DE M tuberculosis fusion protein Tbf-8.
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
XX
OS Synthetic.
OS Mycobacterium tuberculosis.
XX
XX
PN WO9942118-A2.
XX
PD 26-AUG-1999.
XX
XX 17-FEB-1999; 99WO-US003265.
XX
XX 18-FEB-1998; 98US-00024753.
PR
PR 05-MAY-1998; 98US-00072596.
XX
XX (CORI-) CORIXA CORP.
PA
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX
XX
DR WPI; 1999-527416/44.
DR N-PSDB; AA319248.
XX
XX
PT New polypeptide comprising antigenic portions of M. tuberculosis.
XX
XX
PS Example 10; Page 320-321; 323pp; English.
XX
XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against M.
CC tuberculosis infection. The new detection methods are needed as current
CC vaccination strategies do not provide 100% immunity
XX
XX Sequence 652 AA;
SQ

Query Match 86.8%; Score 1499; DB 2; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.5e-89;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPAAPADNAPPPVIAPN 106
Db 374 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPAAPADNAPPPVIAPN 433
Qy 107 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIV 166
Db 434 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIV 493
Qy 167 LGRLDQKLYASAEATDSKAAARLGSMDGGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 226
Db 494 LGRLDQKLYASAEATDSKAAARLGSMDGGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 553
Qy 227 KFSDFSKNGQIWTGVIGSPAANAPDAGPPQRFVWMLGTANNPVDKGAALAESIRPL 286
Db 554 KFSDFSKNGQIWTGVIGSPAANAPDAGPPQRFVWMLGTANNPVDKGAALAESIRPL 613

Qy 287 VAPPPAPAPAPAPAPAPAPAGEVAPTPTTTPQRTLP 325
Db 614 VAPPPAPAPAPAPAPAPAPAGEVAPTPTTTPQRTLP 652

RESULT 15
AA39225
ID AA39225 standard; protein; 652 AA.
XX
XX
AC AA39225;
XX
XX
DT 05-NOV-1999 (first entry)
DE
DE M. tuberculosis fusion protein Tbf-8 amino acid sequence.
XX
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
XX
OS Synthetic.
OS Mycobacterium tuberculosis.
XX
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
XX 17-FEB-1999; 99WO-US003268.
XX
XX 18-FEB-1998; 98US-00025197.
PR
PR 05-MAY-1998; 98US-00072967.
XX
XX (CORI-) CORIXA CORP.
PA
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX
XX
DR WPI; 1999-527409/44.
DR N-PSDB; AA219460.
XX
XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX tests and protective or therapeutic vaccines or compositions.
XX
XX Claim 37; Page 274-276; 299pp; English.
XX
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
CC polypeptides fragments, can be used in pharmaceutical compositions or
CC vaccines to generate a protective or therapeutic immune response to M.
CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
CC to AA219460 and AA39083 to AA39225 are used in the exemplification of
CC the present invention
XX
XX Sequence 652 AA;
SQ

Query Match 86.8%; Score 1499; DB 2; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.5e-89;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPAAPADNAPPPVIAPN 106
Db 374 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPAAPADNAPPPVIAPN 433
Qy 107 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIV 166
Db 434 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIV 493
Qy 167 LGRLDQKLYASAEATDSKAAARLGSMDGGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 226
Db 494 LGRLDQKLYASAEATDSKAAARLGSMDGGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 553

Qy	227	KFSDESKPQNGIWTG	VIGSPAANAPDAGPPQ	RWFVWLGTANNPV	VDKGAAYALAESIRPL	286
Db	554	KFSDESKPQNGIWTG	VIGSPAANAPDAGPPQ	RWFVWLGTANNPV	VDKGAAYALAESIRPL	613
Qy	287	VAPPPAPAPAPAEPA	PAPAPAGEVAPT	TTTTPORTLPA		325
Db	614	VAPPPAPAPAPAEPA	PAPAPAGEVAPT	TTTTPORTLPA		652

Search completed: July 7, 2004, 18:21:03
Job time : 53.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:19:15 ; Search time 22.5 Seconds
(without alignments)
1389.433 Million cell updates/sec

Title: US-10-720-192-2
Perfect score: 1726
Sequence: 1 MHQVDNLTTRKGRLAALAI.....PAGEVAPTPTPTPQRTLPA 325

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1726	100.0	325	2 D70666	probable modD prot
2	980	56.8	287	2 B87166	hypothetical prote
3	241	14.0	1585	2 T31611	hypothetical prote
4	237.5	13.8	1188	2 S49915	extensin-like prot
5	228.5	13.2	464	2 S22697	extensin - Volvox
6	226.5	13.1	222	2 H96711	hypothetical prote
7	224.5	13.0	3534	2 T42567	tegument protein 2
8	219.5	12.7	85	2 T44744	antigen 43L [impor
9	217.5	12.6	817	2 S51342	verprolin - yeast
10	217.5	12.6	839	2 F75518	hypothetical prote
11	217	12.6	801	2 T29018	hypothetical prote
12	216.5	12.5	847	2 F96531	hypothetical prote
13	213.5	12.4	760	2 T06291	extensin homolog T
14	213.5	12.4	3164	1 WWB8H5	UL36 protein - hum
15	208.5	12.1	214	2 T10737	extensin-like cell
16	207	12.0	416	1 SXXLAG	dermal gland prote
17	206.5	12.0	1611	2 T28236	hypothetical prote
18	204.5	11.8	539	2 T28770	hypothetical prote
19	203	11.8	1201	2 G86441	unknown protein [i
20	203	11.8	3421	1 W2BEB6	367K tegument prot
21	202.5	11.7	283	2 S13383	hydroxyproline-ric
22	201.5	11.7	395	2 H75457	hypothetical prote
23	200.5	11.6	241	2 T09854	proline-rich cell
24	200.5	11.6	760	2 F86387	probable Pto kinas
25	199	11.5	708	2 D96711	hypothetical prote
26	199	11.5	839	2 T04859	extensin homolog F
27	197.5	11.4	865	2 A47282	calcium-binding pr
28	196.5	11.4	418	2 T15142	hypothetical prote
29	196.5	11.4	666	2 B70803	hypothetical prote

30	194	11.2	191	2 F84522	probable proline-r
31	193.5	11.2	1151	2 T18535	high molecular mas
32	193	11.2	620	2 S06733	hydroxyproline-ric
33	192.5	11.2	731	2 T04455	hypothetical prote
34	192.5	11.2	744	2 E86255	hypothetical prote
35	192.5	11.2	873	2 A47283	calphostin-like frui
36	191.5	11.1	280	2 T11671	extensin-like prot
37	191	11.1	576	2 T36729	probable serine/th
38	190	11.1	550	2 G70597	probable proteinas
39	189.5	11.0	267	2 S08314	cell wall glycopro
40	189	11.0	240	2 B24264	proline-rich prote
41	189	11.0	1794	2 T38459	hypothetical diver
42	188	10.9	544	2 T17547	proline-rich prote
43	188	10.9	699	2 C43674	US4 protein - huma
44	188	10.9	1268	2 T31420	C-terminal domain-
45	187	10.8	350	2 S22456	hydroxyproline-ric

ALIGNMENTS

RESULT 1

D70666

probable modD protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: D70666

R;Cole, S.T.; Brogi, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtrovd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537(544), 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70360; PMID:98295987; PMID:9634230

A;Accession: D70666

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-325 <COL>

A;Cross-references: GB:Z83859; GB:AL123456; NID:G3261678; PIDN:CAB06127.1; PID:e290722,1

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: modD

Query Match	100.0%	Score 1726;	DB 2;	Length 325;
Best Local Similarity	100.0%	Pred. No. 2.8e-87;		
Matches 325;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 MHQVDNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPAPVPVPTTAASTPSTAAA 60

Db 1 MHQVDNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPAPVPVPTTAASTPSTAAA 60

Qy 61 PPAPATPVAPPPPPAAANTPNAOPGDPNAPPPVIAPNAPOPVIDNPVGGF 120

Db 61 PPAPATPVAPPPPPAAANTPNAOPGDPNAPPPVIAPNAPOPVIDNPVGGF 120

Qy 121 SFALPAGWESDAAHFDYGSALLSKTTGDPFPFGGPPPVANDTRIVLGRDQKLASABA 180

Db 121 SFALPAGWESDAAHFDYGSALLSKTTGDPFPFGGPPPVANDTRIVLGRDQKLASABA 180

Qy 181 TDSKAARLGSDMGGEYMPYPGTRINQETVSDANGVSGSASVYEVKFSDDPSKNGQIWT 240

Db 181 TDSKAARLGSDMGGEYMPYPGTRINQETVSDANGVSGSASVYEVKFSDDPSKNGQIWT 240

Qy 241 GVIGSPAANAPDAGPPQRFVFWLGTANNPVDKGAALAESIRPLVAPPPAPAPAP 300

Db 241 GVIGSPAANAPDAGPPQRFVFWLGTANNPVDKGAALAESIRPLVAPPPAPAPAP 300

Qy 301 APAPAPAGEVAPTPTPTPTQRTLPA 325

Db 301 APAPAPAGEVAPTPTPTPTQRTLPA 325

RESULT 2

C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
 C;Accession: S22697; S21006
 R;Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
 EMBO J. 11, 2055-2062, 1992
 A;Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox carterii
 A;Reference number: S22697; MUID:92289669; PMID:1600938
 A;Accession: S22697
 A;Molecule type: mRNA
 A;Residues: 1-464 <HL>
 A;Cross-references: EMBL:X65165; NID:G21991; PIDN:CAA46283.1; PID:G21992
 C;Keywords: Glycoprotein

Query Match 13.2%; Score 228.5; DB 2; Length 464;
 Best Local Similarity 22.3%; Pred. No. 1e-05;
 Matches 70; Conservative 16; Mismatches 69; Indels 157; Gaps 9;

QY 34 PATANADPEA-----PPVPTTAASPSSTAAAPAPATVA--PPP 72
 DB 235 PARVSSSPATRRPPRRITSPVLTASPLPKT--SPFPRVPPSPPPVAGSPPPP 292
 QY 73 PAAATNAQGDNDNAAPPADP---NAPPPVIAADNAPQVRIIDNVPVGGFALPAGW 128
 DB 293 PPRVSPSPPPQVSPSPPPPPRPSPPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 334
 QY 129 VESDAAHFYGSALLSKTTGDPFPQPPVANDTRIVLGRLDQKLYASAEATDSKAAAR 188
 DB 335 -----SPPPSP 349
 QY 189 LGSDMGFYMPYPCTRINQETVSLDANGVSGSASYEVKPSDKPQGIWTCVIGSPAA 248
 DB 350 -----PPPPR-----SSPSPPPPV-----SPPP 368
 QY 249 NADPAGPQRFVVMVWLTANNPVDKGAALAEBSIRPLVAPPAPAPAPAPAPAPAG 308
 DB 369 PPRASPPP-----PPASPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 402
 QY 309 EVAPTPPTTPTPQRT 322
 DB 403 TAAANPSPAPSR 416

RESULT 6
 H96711
 hypothetical protein F14K14.17 [imported] - Arabidopsis thaliana
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C;Accession: H96711
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A;Reference number: A86141; MUID:21016719; PMID:11130712
 A;Accession: H96711
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-222 <STO>
 A;Cross-references: GB:AB005173; NID:G6524187; PIDN:AAF15072.1; GSPDB:GN00141
 C;Genetics:
 A;Gene: F14K14.17
 A;Map position: 1

Query Match 13.1%; Score 226.5; DB 2; Length 222;
 Best Local Similarity 23.1%; Pred. No. 6.2e-06;
 Matches 75; Conservative 14; Mismatches 75; Indels 161; Gaps 9;

QY 22 AMASASLTVAVPATANADPPAPVPTT-----AASPPSTAAAPAPATPVAPPAAA 75

DB 28 AASPVTSTTTAPPPTTAAPPTTAAPPTTTTPPVSAQQPASPVTTPPPPAVTTSPAPKV 87
 QY 76 AN--TPNAQPGDNAAAPPADENAPPPVIAADNAPQVRIIDNVPVGGFALPAGVESA 133
 DB 88 ADVISPATPPQPPSPSPASAPTTSPPVSPPPAP----- 122
 QY 134 AHFDYGSALLSKTTGDPFPQPPVANDTRIVLGRLDQKLYASAEATDSKAAARLGSMD 193
 DB 123 -----TSPPPTTASPP-----APASPPPPAPA----- 144
 QY 194 GFYMPYPCTRINQETVSLDANGVSGSASYEVKPSDKPQGIWTCVIGSPAANAPDA 253
 DB 145 -----SPPPAP-----VSPPPVQAPSP 161
 QY 254 GPPQRFVVMVWLTANNPVDKGAALAEBSIRPLVAPPAPAPAPAE-----P 300
 DB 162 -----ISLPAPAPAPTKHKRKHKHHHAP 188

RESULT 7
 T42567
 tegument protein 24 - equine herpesvirus 4 (strain NS80567)
 C;Species: equine herpesvirus 4
 A;Variety: strain NS80567
 C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C;Accession: T42567
 R;Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
 J. Gen. Virol. 79, 1197-1203, 1998
 A;Title: The DNA sequence of equine herpesvirus-4.
 A;Reference number: Z22173; MUID:98264497; PMID:9603335
 A;Accession: T42567
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-3534 <TEL>
 A;Cross-references: EMBL:AF030027; NID:G2605950; PIDN:AAC59539.1; PID:G2605967
 A;Experimental source: strain NS80567
 C;Genetics:
 A;Note: 24
 C;Superfamily: varicella-zoster virus gene 22 protein

Query Match 13.0%; Score 224.5; DB 2; Length 3534;
 Best Local Similarity 27.8%; Pred. No. 0.00012;
 Matches 91; Conservative 18; Mismatches 93; Indels 125; Gaps 13;

QY 5 DPNLTRRGRGLAALAAMASASLTVAVPATANADPPAPVPTTAASP--PSTAAAPP 62
 DB 2710 DPN-----EALLTAPSKPAAAPAPSKPAAAPAPSKPAAAPAPSKPAAAPAPSK 2765
 QY 63 APATPVAPPPPPPAAANTPNAPQDPNAPPPADPNAPPPVIAADNAPQVRIIDNVPVGGFSF 122
 DB 2766 KPAAPAPSKPAAAPAPSK-----KPAAPAPSKPAAAPAPSKPAAAPAPSK----- 2811
 QY 123 ALPAGVESDAAHFYGSALLSKTTGDPFPQPPVANDTRIVLGRLDQKLYASAEATD 182
 DB 2812 ---PAA-----APAPSKP-----AAAPAPSK 2828
 QY 183 SKAAARLGSMDMGFYMPYPCTRINQETVSLDANGVSGSASYEVKPSDKPQGIWTCV 242
 DB 2829 KPAAP-----PAPSKP-----PAPSKP----- 2839
 QY 243 IGSPAANADAGP-PQRFVVMVWLTANNPVDKGAALAEBSIRPLVAP---PPAPAPAPA 298
 DB 2840 AAAPAPSKPAAAPAPSK-----PAAAPAPSKPAAAPAPSKPAAAPAPSKPAAAPAPSK 2891
 QY 299 EPAPAPAPAGEVAPTPTTPTPQRTLPA 325
 DB 2892 KPAAPAPSKPAAAPAPSKPQNTLVA 2917

RESULT 8

```

T44744      antigen 43L [imported] - Mycobacterium leprae (fragment)
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T44744
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
   submitted to the EMBL Data Library, October 1997
A:Reference number: Z2833
A:Accession: T44744
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-85 <PAR>
A:Cross-references: EMBL:AL008609; PIDN:CAA15433.1
A:Experimental source: cosmid B1788
C:Genetics:
A:Note: 43L

Query Match      12.7%; Score 219.5; DB 2; Length 85;
Best Local Similarity 52.1%; Pred. No. 5.7e-06;
Matches 50; Conservative 8; Mismatches 27; Indels 11; Gaps 2;

Qy      1  MQQVDPNLTFRKGLAALAIAAMASASLTVAVPATANADPEPAPVPTTAASPPSTAAA 60
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1  MQQVLDLSTHRKGLWAILAIIVVASASATMELPAANADDPAPL-----PPSTATA 51

Qy      61  PPAPATPVAPPPPPAAANTPNAQPGDPNAPPPADPN 96
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      52  APSFAQEIITPLPGAPVSSEAGPGDPNA--PSLDPN 85
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9

S51342
 verprolin - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: prolin-rich protein VRP1; protein L8300.l3; protein YLR337c
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 23-Feb-1995 #sequence revision 11-Aug-1995 #text_change 23-Mar-2001
 C:Accession: S51342; S39626; S57435
 R:Du, Z.
 submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of *S. cerevisiae* cosmid 8300.
 A:Reference number: S51339
 A:Accession: S51342
 A:Molecule type: DNA
 A:Residues: 1-817 <DUZ>
 A:Cross-references: EMBL:U19028; NID:G609380; PID:G609392; MIPS:YLR337c
 R:Donnelly, S.F.H.; Pocklington, M.J.; Pallotta, D.; Orr, E.
 Mol. Microbiol. 10, 585-596, 1993
 A:Title: A proline-rich protein, verprolin, involved in cytoskeletal organization and cell growth in *Saccharomyces cerevisiae*
 A:Reference number: S39626; MUID:95058201; PMID:7968536
 A:Accession: S39626
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-307,'R',309-349,'R',351-688,'E',690-709,'HLRWIPVELIAPVKTLNNGYFLQVDRRCNTSII
 A:Cross-references: EMBL:Z26645; NID:G414785; PIDN:CAA81388.1; PID:G439289
 R:Munn, A.L.; Stevenson, B.J.; Geli, M.I.; Riezman, H.
 submitted to the EMBL Data Library, June 1995
 A:Description: ends, end6 and end7: mutations that cause actin delocalization and block
 A:Reference number: S57435
 A:Accession: S57435
 A:Molecule type: DNA
 A:Residues: 1-162,'F',164-817 <MDN>
 A:Cross-references: EMBL:X87806; NID:G871534; PID:G871535
 R:Experimental source: strain W303

F;349-357/Region:	proline-rich
F;372-382/Region:	proline-rich
F;396-406/Region:	proline-rich
F;421-445/Region:	proline-rich
F;518-528/Region:	proline-rich
F;567-577/Region:	proline-rich
F;608-621/Region:	proline-rich
F;649-661/Region:	proline-rich
F;678-685/Region:	proline-rich
F;704-710/Region:	proline-rich

Query Match 12.6%; Score 217.5; DB 2; Length 817;
Best Local Similarity 26.3%; Pred.No.6.9e-05;
Matches 99; Conservative 31; Mismatches 125; Indels 121; Gaps 18;

QY	24 ASASLVTVAVPATA-----NADPEFAPPVPT--TAASPPP-----STAAAPPAATPVAFPPP 72
DB	118 ASAPIPGAVPSVAAPPIENAPLSAPAAPSIPSSSAPPIDIPSRAAIIPIIVSPASP 177
QY	73 P-----AAANTPNAGCPDMAAP-----PPA-----DPN 96
DB	178 PLPLSGASAPKYPQNRRHPMSVRFAHRSHQRKSNIILSPSVSAIPLPSALPTHVSNNPPQ 237
QY	97 APPRPVV-----TAP--NAQPVRINDNPVGFGFSFALPWGTVESDAAHFDYGS---- 140
DB	238 A PPPPPTTGIGDSKNIKPTONANSPSSSEVPACGLFL-----AEINARSRERGAVEGV 292
QY	141 -----ALLSKTTGDPPFPQGPFPPVAND-----TRVLGRLDOKUYASAENAETDKAAA 187
DB	293 SSTKIQTENHKSPSQPLPSSAPPITPSHAPLPPTAPPSPSLPNVTSAKKATSADAPP 352
QY	188 RLGGDMGEFFMYPYCRTNRINOETVSLDANGVSGSYVEVKFSDPKXNGQIWTCVIGSPA 247
DB	353 P-----PPLPAAMSSASTNSVKATPVPTTLIA-----PPLEN---TTSVPPNKKA 392
QY	248 ANPADGAPQQRFVFVWLGTANNPDVKGA---AKALAERSIRPLVAPPAPPAEAPAP-A 303
DB	393 SSMEAPPPPP-----PPPGAFTSSALSASSIFLAFLPPPPPSVATSVPSA 440
QY	304 PAPAGEVAPTPTTTP 319
DB	441 PPP-----PPTLTINKP 452

RESULT 10

```

F75518
hypothetical protein - Deinococcus radiodurans (strain R1)
C Species: Deinococcus radiodurans
C Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C Accession: F75518
R White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Fraser, C.M.; Shen, M.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mahoney, S.; Smith, H.O.; Waterman, W.H.; Davis, K.P.; Adams, M.D.; McQuinn, K.L.; et al., Science 286, 1571-1577, 1999
A Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A Reference number: A75250; UID:20036896; PMID:10567266
A Accession: F75518
A Status: Preliminary
A Molecule type: DNA
A Residues: 1-839 <WH>
A Cross-references: GB:AEO001904; GB:AEO00513; NID:g6458129; PID:NAAF10038.1; PID:g6458141; Experimental source: strain R1
    
```

QY
| | | | |
22 AMASASLTVAVPATANAD-----PEPAP--PVPTTAASPSTAAAP----PAPATVPAP- 70

Query Match 12.6%; Score 217.5; DB 2; Length 839;
Best Local Similarity 25.4%; Pred. NO. 7.1e-05;
Matches 95; Conservative 24; Mismatches 124; Indels 131; Gaps 17;

C/Superfamily: herpesvirus immediate-early protein IEL0; RING finger homology

A/Gene position: 1

A/Gene: D0108

Query Match.	12.6%;	Score 217;	DB 2;	Length 801;
Best Local Similarity	24.9%;	Pred. No. 7.2e-05;		
Matches 101; Conservative	37;	Mismatches 151;	Indels 116;	Gaps 16;
Qy	25 SASLVTWAVZATANADPEAP-----FVPTTAAS---PPST-----	57		
	:: :: : :			
Db	194 ASGVYTSQQGAGDAEPFPVVVEETPAPTAAEETPAPAETAASNAVGTPEGYYDGTANA	253		
Qy	58 --AAAAPATPATPVAPPAAANTP-NAQPGDPNAAPPADPNAPPPVPIAFNAPQPVRIID	114		
	:: :: : :			
Db	254 ASAVAEEFAVETPA--PAPAAEETPAPATSSEEAAPAPATAEETPAPTETVSA-APEAANSY	310		
Qy	115 NPVGFGFALLPWGES-----DAAHFDYGSAKLSKITGD-----	149		
	:: :: : :			
Db	311 DSAGGDAATAPAPSSEAADAAPDTDSASAADTTTAAALVDTSSEHAESTAPATDDIAATE	370		
Qy	150 -PPPGGGPPVAN-----DIIVLGRLDOKLYASAEATDKSAAARGLSGDMGEF	196		
	:: :: : :			
Db	371 TTPAPSPVPADVADAAGDYGSPIETTPAAPAEDTPAPASAAAEETPAPAAEE----	426		
Qy	197 YMPYPGTR-----INQETVSUDANGVSGSAYYEKVTFSDPSKPNGQI	238		
Db	427 -TZAPETASAADGAACCAAPADVAAPADVATTAPETSQAAGSYDV----PSEPASEV	481		

Query Match	12.5%	Score	216.5	DB 2	Length	847			
Best Local Similarity	26.0%	Pred. No.	8.1e-05						
Matches	79	Conservative	35	Mismatches	135	Indels	55	Gaps	12
Qy	33	VPATANADRPAPVPTTAASBPSTAAAPAPATPVAP	---	PPAAANTENAQPGDPNA	88				
Db	533	VPDPQPMPSPPSPPIYSPPPVHSGPPPVYSGPPPHVYSPPPPVVASPPPPSPPPVH	592						
Qy	89	APPPADPNAPPPVPIAFNAPQPVRIQNPVGGF3FALPAGWVESDAAHFDYGSALLSKTTG	148						
Db	593	SPPPPPPVFSPPPVFSPPPPSPVYSPPPP	---	SHSPP	---	---	---	PPVYS	631
Qy	149	DPFPFGQGPVANDTRIVLG	-	RLDQKLYASAAATUSKAAARLGSNDGCFYMPYFG	-	TRI	205		
Db	632	PPPTTSPPTTHNTNPPMGPAPTQTOAPTSPSETTQVPTPSS	-	ESDOSQILSPVQAPTPV	690				
Qy	206	NQTVSLDANGV	---	SGSASYEYKVSDDPSKPNQIWTGVIGSPAANAPAGPQPQRFV	262				
Db	691	QSSTPSPSEPTQVTPSPSSSYQAPNUSPVQAPT	---	PVQAPTTSSETTSQVP	---	---	738		
Qy	263	WLGTANNPVDKGAALKAAESIRPLVAPPP	---	APAPAPAPAPAPAGEV	-	APTPT	315		
Db	739	---	TPSSESNQSPQAPTPPLEFVHAPTPNSKPVQSPTPS	-	SFVSSPSEQSEVEAEPT	794			
Qy	316	TPPT	319						
Db	795	VPNP	798						

RESULT 13
T06291
extensin
C;Species
C;Date: 3
C;Accessi
R;Bevan,
submitted
A;Referen

RESULT 13

T06291

extensin homolog T988.80 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 30-Apr-1999 #sequence revision 30-Apr-1999 #text_change 11-Jun-1999
C/Accession: T06291
R/Bavan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Newes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, April 1999
A/Reference number: Z1558

A:Accession: T06291
A:Molecule type: DNA
A:Residues: 1-760 <BEV>
A:Cross-references: EMBL:AL049608
A:Experimental source: cultivar Columbia; BAC clone T958
C:Genetics:
A:Map position: 4
A:Note: T9E8.80

Query Match 12.4%; Score 213.5; DB 2; Length 760;
Best Local Similarity 23.6%; Pred. No. 0.00011; Indels 119; Gaps 13;
Matches 78; Conservative 26; Mismatches 107;

QY 28 LVTAVPATANADPEAP--PVPTTAASPPSTAAPAPATPVAPPVPPAAANTPNAQPGD 85
DB VVTLPPESLSPPPPAFIFSTPTLTSPPPSPPPVSPPPPPPPPPVSPPPPPPP 457
QY 86 P-----NAAPPADPNAPPPVIAPNAPQVRIQNVGGSFALPAGWESDAAHFDYGS 140
DB PPPPPVYSPPPPPPPPPPPVYSPPPSPPPPPV-----YSP----- 498
QY 141 ALLSKTTGDPGPPGPPVANDTRIVLGRDQKLYASAEATDSKAAARL----- 189
DB -----PPPPPPPPVYSP-----PPVYSSPPPPSPAPTVYCTRPVPPPH 543
QY 190 GSDMGFEYMPYGTINQETVSLDANGVSGSASYEYVKFSDPKNGQIWTGIVGSPAAN 249
DB 544 SPPPPQFSPPP-----EPYV-----YSSPPPPH-----SSPPPH 573
QY 250 APDAGPPQRFVWLGTTANNVVDKGAALAESIRPLVAPPAPAPAPAP----- 300
DB 574 SP-----PP-----PHSP-----PPPIYPLSPPPPPPTVSSPPPTPVYSPPP 611
QY 301 -----APAPAGVAPVPTTPTPQRTL 324
DB 612 PPPCTEPPPPPCIEYSPPPPPVHYSSP 641

RESULT 14
WMBEH6
UL36 protein - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jun-2000
C:Accession: I30085
R:McGeoch, D.J.; Dairymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perr
J. Gen. Virol. 69, 1531-1574, 1988
A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim
A:Reference number: A30083; MUID:88274327; PMID:2839594
A:Accession: I30085
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-3164 <MCG>
A:Cross-references: GB:X14112; NID:G1944536; PIDN:CAA32311.1; PID:G59536; GB:D00317
C:Genetics:
A:Gene: UL36
C:Superfamily: varicella-zoster virus gene 22 protein

Query Match 12.4%; Score 213.5; DB 1; Length 3164;
Best Local Similarity 26.5%; Pred. No. 0.00044; Indels 119; Gaps 15;
Matches 91; Conservative 17; Mismatches 117;

QY 10 RKGRALALAIAMASA-----SLVTAVPATANADPEAPVPTTAAASPESTAAAP 61
DB 2683 RRRREARSLSATQASATTCWRPPALPDVAVPTDFARP-PAPKPPPEAPHALVSGVP 2741
QY 62 -----PAPATPVAP-PPPAANTPNAQGDPAAPPPADFNAPPPVIAPNAPQ 110
DB 2742 LPLGFQAAGQASPALPIDVPPPPVATGT--VLPGENRRRPLTSGPAPTTPRVVGGPQR 2799
QY 111 VRIDNPVGGFSFALPAGWESDAAHFDYGSALLSKTTGDPGPPGPPVANDTRIVLGR 170
DB 2800 -RLTFP-----AVASLSRESLSPWDP---ADPTAPVLR- 2832

QY 171 DQKLYASAEATDSKAAARLGSDMGFBYMPYPCTRINQETVSLDANGVSGSASYEYVKFSD 230
DB 2833 -----NPAPTSSSPAG----- 2844
QY 231 PSKNGQIWTGIVGSPAANADPAGPPQRFVFWLGT-----NNPVDKGAALAESIR 284
DB 2845 PSPPPPAV-----QFVAPPTTSGPPTTLTLEGVAPGPGFVSRRPTTRQPVATFTTSAR 2898
QY 285 P-----LVAPPAPAPAPAPAPAPAGVAPTPTTTPQ 320
DB 2899 PRGHLTVSLSPAPQPPQPP-PPQPPQPPQPPQPP-PPQPPQ 2940

RESULT 15
T10737
extensin-like cell wall protein - sea-island cotton
N:Alternate names: proline-rich cell wall protein
C:Species: Gossypium barbadense (sea-island cotton)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10737
R:John, M.E.; Keller, G.
Plant Physiol. 108, 669-676, 1995
A:Title: Characterization of mRNA for a proline-rich protein of cotton fiber.
A:Reference number: Z16888; MUID:95334493; PMID:7610164
A:Accession: T10737
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-214 <JOH>
A:Cross-references: EMBL:U04367; NID:G451543; PIDN:AAA79364.1; PID:G451544
A:Experimental source: fiber
C:Genetics:
A:Introns: 187/3
C:Superfamily: proline-rich protein
C:Keywords: cell wall

Query Match 12.1%; Score 208.5; DB 2; Length 214;
Best Local Similarity 25.1%; Pred. No. 5.7e-05;
Matches 76; Conservative 17; Mismatches 73; Indels 137; Gaps 13;

QY 15 LAALAIAAMASASLVTAVPATANADPEAPVPTTAAASPPSTAAPAPATPVAPP 74
DB 10 LSALLCIAVAGVLGQAPSNPTST----PAPPTP-PASTPPTTQAPP---TPTATPPPV 61
QY 75 AANTNAQGDENAAAPPPAD---PNAPPPVIAPNAPQVRIQNVGGSFALPAGWES 131
DB 62 STPPPTSSPPVTAAGPPVSTTPPSPPTPFPASPP----- 100
QY 132 DAAHFDYGSALLSKTTGDPGPPGPPVANDTRIVLGRDQKLYASAEATDSKAAARLGS 191
DB 101 -----ATPP-PASPPP-----ATPPAS----- 117
QY 192 DMGEFVMPYGTINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWTGIVGSPAANAP 251
DB 118 -----PPFAT-----PPPATP-----PPATPP 134
QY 252 DAGPPQRFVFWLGTANNPVDKGAALAESIRPLVAP---PPAP---APAPAEFAPAPA 305
DB 135 PATPP-----PAPLASPPATVPALSPVQTLTSPFPAPTTEAPTLGAATPG 181
QY 306 PAG 308
DB 182 PAG 184

Search completed: July 7, 2004, 18:24:00
Job time : 25.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:17:39 ; Search time 10.5 Seconds

(without alignments)
1611.695 Million cell updates/sec

Title: US-10-720-192-2

Perfect score: 1726

Sequence: 1 MHQVDPNLTTRKGRKLAALAI.....PAGEVAPTPTPTPTQTLPA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1726	100.0	325	1 APA_MYCTU	Q50906 mycobacteri
2	1720	99.7	325	1 APA_MYCBO	Q30620 mycobacteri
3	985	57.1	381	1 APA_MYCAV	Q48919 mycobacteri
4	980	56.8	287	1 APA_MYCLE	P45842 mycobacteri
5	243	14.1	555	1 GPI_CHLRE	Q95PQ6 chlamydomon
6	217.5	12.6	817	1 VRP1_YEAST	P37370 saccharomyc
7	213.5	12.4	3164	1 TEGU_HSV11	P10220 herpes simp
8	210	12.2	1083	1 T2D3_HUMAN	O00268 homo sapien
9	207	12.0	439	1 XP2_XENLA	P17437 xenopus lae
10	205.5	11.9	802	1 ENAH_MOUSE	Q03173 mus musculu
11	203	11.8	3421	1 TEGU_HSVB	P28955 equine herp
12	197.5	11.4	865	1 CPN_DROME	Q02910 drosophila
13	197	11.4	283	1 EXTN_SORBI	P24152 sorghum bic
14	193	11.2	620	1 EXTN_TOBAC	P13983 nicotiana t
15	192.5	11.2	1198	1 HCN4_RAT	Q9JKA7 rattus norv
16	191	11.1	1386	1 ZAP3_MOUSE	Q9R017 mus musculu
17	189.5	11.0	267	1 EXTN_MAZE	P14918 zea mays (m
18	189	11.0	1794	1 YDC9_SCHPO	Q10172 schizosacch
19	189	11.0	2167	1 SHK1_RAT	Q9WV48 rattus norv
20	188.5	10.9	2715	1 MLL4_HUMAN	Q9UMN6 homo sapien
21	188	10.9	699	1 VGLG_HSV2H	P13290 herpes simp
22	186.5	10.8	5179	1 MUC2_HUMAN	Q02817 homo sapien
23	186	10.8	331	1 PRP1_HUMAN	P04280 homo sapien
24	186	10.8	1033	1 IF2_STRCO	Q8CJQ8 streptomyc
25	185.5	10.7	1300	1 SAL3_HUMAN	Q9BXA9 homo sapien
26	184.5	10.7	639	1 ZIC5_HUMAN	Q96T25 homo sapien
27	184.5	10.7	1046	1 IF2_STRAW	Q82K53 streptomyc
28	184	10.7	1271	1 RBMG_HUMAN	Q9UPN6 homo sapien
29	183	10.6	347	1 CSP_FLABA	P23093 plasmodium
30	182	10.5	261	1 PRP2_MOUSE	P05142 mus musculu
31	180.5	10.5	534	1 APG_ARATH	P40602 arabidopsis
32	180.5	10.4	2716	1 OSA_DROME	Q8IN94 drosophila
33	179.5	10.4	2142	1 BAT2_HUMAN	P48634 homo sapien

ALIGNMENTS

RESULT 1

APA_MYCTU
ID APA_MYCTU STANDARD; PRT; 325 AA.
AC Q50906; O08062;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alanine and proline-rich secreted protein apa precursor (45/47 kDa antigen) (Fibronectin attachment protein) (Immunogenic protein MPT32)
DE (Antigen MPT-32) (45-kDa glycoprotein) (FAP-B).
GN APA OR MODD OR RV1860 OR MT1908 OR MTCY359.13.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID:1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=96009758; PubMed=7558311;
RA Laqueyrie A., Miltzer P., Romain F., Eiglmeier K., Cole S., Marchel G.;
RT Cloning, sequencing, and expression of the apa gene coding for the Mycobacterium tuberculosis 45/47-kilodalton secreted antigen complex. ;
RL Infect. Immun. 63:4003-4010(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX Laqueyrie A.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Seesch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.B. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sultston J., Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. ;
RL Nature 393:537-544(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551; Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Klenz J.A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. ;

RA J. Bacteriol. 184:5479-5490(2002).
 RN [5] PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE=9109989; PubMed=1898899;
 RA Nagai S., Wiker H.G., Harboe M., Kinomoto M.;
 RT "Isolation and partial characterization of major protein antigens in
 RL the culture fluid of Mycobacterium tuberculosis.";
 RN Infect. Immun. 59:372-382(1991).
 RA [6] PARTIAL SEQUENCE, AND GLYCOSYLATION.
 RX MEDLINE=95347792; PubMed=7622204;
 RA Dobos K.M., Swiderek K., Khoo K.-H., Brennan P.J., Belisle J.T.;
 RT "Evidence for glycosylation sites on the 45-kilodalton glycoprotein of
 RL Mycobacterium tuberculosis.";
 RN Infect. Immun. 63:2846-2853(1995).
 RA [7] PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=96196153; PubMed=8626314;
 RA Dobos K.M., Khoo K.-H., Swiderek K.M., Brennan P.J., Belisle J.T.;
 RT "Definition of the full extent of glycosylation of the 45-kilodalton
 RL glycoprotein of Mycobacterium tuberculosis.";
 RN J. Bacteriol. 178:2498-2506(1996).
 RA [8] SEQUENCE OF 40-49, AND CHARACTERIZATION OF CARBOHYDRATE-LINKAGE SITES
 BY MASS SPECTROMETRY.
 RX MEDLINE=20011399; PubMed=10542234;
 RA Horn C., Namane A., Pescher P., Riviere M., Romain F., Puzo G.,
 Barzu O., Marchal G.;
 RT "Decreased capacity of recombinant 45/47-kDa molecules (Apa) of
 RL Mycobacterium tuberculosis to stimulate T lymphocyte responses
 RT related to changes in their mannosylation pattern.";
 RN J. Biol. Chem. 274:32023-32030(1999).
 RA [9] SEQUENCE OF 40-57, AND CHARACTERIZATION OF CARBOHYDRATE-LINKAGE SITES
 BY MASS SPECTROMETRY.
 RX MEDLINE=20002527; PubMed=10531201;
 RA Romain F., Horn C., Pescher P., Namane A., Riviere M., Puzo G.,
 Barzu O., Marchal G.;
 RT "Deglycosylation of the 45/47-kilodalton antigen complex of
 RL Mycobacterium tuberculosis decreases its capacity to elicit in vivo
 RT or in vitro cellular immune responses.";
 RN Infect. Immun. 67:5567-5572(1999).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MASS SPECTROMETRY: MW=28782; MW_ERR=1.6; METHOD=Electrospray;
 CC RANGE=40-325.
 CC -!- BIOTECHNOLOGY: Major immunodominant antigen that has potential as
 CC a vaccine against tuberculosis. APA-ELISA could be used in
 CC diagnosis.
 CC -!- MISCELLANEOUS: Changes in the mannosylation pattern of this
 CC protein affect its ability to stimulate T-lymphocyte response.
 CC -!- CAUTION: Was originally thought to be involved in molybdenum
 CC transport.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X80268; CAA56555.1; -
 DR EMBL; X99258; CAA67645.1; -
 DR EMBL; Z83859; CAB06127.1; -
 DR EMBL; AB007048; AAK46179.1; -
 DR PIR; D70666; D70666.
 DR TIGR; MT1908; -
 DR Trunculist; Rv1860; -
 RA Antigen; Glycoprotein; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 39
 FT CHAIN 40 325 ALANINE AND PROLINE-RICH SECRETED PROTEIN

FT DOMAIN 85 107 APA.
 FT 3 X 4 AA APPROXIMATE REPEATS OF [DA]-P-
 FT N-A.
 FT REPEAT 85 88 1.
 FT REPEAT 94 97 2.
 FT REPEAT 104 107 3.
 FT CARBOHYD 49 49 O-LINKED (MAN. . .).
 FT CARBOHYD 57 57 O-LINKED (MAN. . .).
 FT CARBOHYD 66 66 O-LINKED (MAN. . .).
 FT CARBOHYD 316 316 O-LINKED (MAN. . .).
 FT CONFLICT 136 136 F -> L (IN REF. 4).
 SQ SEQUENCE 325 AA; 32720 MW; 59B5D0455A997BED CRC64;
 Query Match 100.0%; Score 1726; DB 1; Length 325;
 Best Local Similarity 100.0%; Pred. No. 1.5e-80;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEQVDFNLTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPVPTTAAASPESTAAA 60
 DB 1 MEQVDFNLTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPVPTTAAASPESTAAA 60
 QY 61 PPAPATPVAPPPAAANTPNAQPGDPNAPPPADPNAPPPVIAAPNAPQPVRIIDNPVGGF 120
 DB 61 PPAPATPVAPPPAAANTPNAQPGDPNAPPPADPNAPPPVIAAPNAPQPVRIIDNPVGGF 120
 QY 121 SPALPAGWVESDAAHFDYGSALLSKTTGDPGGPPGPPVANDTRIVLGRLOKLYASAEA 180
 DB 121 SPALPAGWVESDAAHFDYGSALLSKTTGDPGGPPGPPVANDTRIVLGRLOKLYASAEA 180
 QY 181 TTSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSYYEVKTSKPNQIWT 240
 DB 181 TTSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSYYEVKTSKPNQIWT 240
 QY 241 GVIGSPAANADAGPPQRFVVMVLTANNPVDKGAALAEISIRPLVAPPAPAPAPAP 300
 DB 241 GVIGSPAANADAGPPQRFVVMVLTANNPVDKGAALAEISIRPLVAPPAPAPAPAP 300
 QY 301 APAPAGEVAPT 325
 DB 301 APAPAGEVAPT 325
 RESULT 2
 APA_MYCBO
 ID APA_MYCBO STANDARD; PRT; 325 AA.
 AC G30620;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Alanine and proline-rich secreted protein apa precursor (Fibronectin
 DE attachment protein) (45/47 kDa antigen) (FAP-B).
 GN APA OR ME1891.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BCG;
 RA Zhao W., Schorey J.S., Bong-Mastek M., Brown E.J., Ratliff T.L.;
 RT "Identification, sequence and characterization of the M. bovis BCG
 RT fibronectin attachment protein.";
 RN Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RN Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

CC -!- SUBCELLULAR LOCATION: Secreted.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF013569; AAB71842.1; -;
 DR EMBL; BX248340; CAD94594.1; -;
 KW Antigen; Glycoprotein; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 39 BY SIMILARITY.
 FT CHAIN 40 325 ALANINE AND PROLINE-RICH SECRETED PROTEIN
 FT CHAIN 40 325 APA.
 FT DOMAIN 85 107 3 X 4 AA APPROXIMATE REPEATS OF [DA]-P-
 FT REPEAT 85 88 N-A.
 FT REPEAT 85 88 1.
 FT REPEAT 94 97 2.
 FT REPEAT 104 107 3.
 FT CARBOHYD 49 49 O-LINKED (MAN. . .) (BY SIMILARITY).
 FT CARBOHYD 57 57 O-LINKED (MAN. . .) (BY SIMILARITY).
 FT CARBOHYD 66 66 O-LINKED (MAN. . .) (BY SIMILARITY).
 FT CARBOHYD 316 316 O-LINKED (MAN. . .) (BY SIMILARITY).
 SQ SEQUENCE 325 AA; 32686 MW; D3419CA5547D91E9 CRC64;

Query Match 99.7%; Score 1720; DB 1; Length 325;
 Best Local Similarity 99.7%; Pred. No. 3e-80;
 Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHQVDNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
 DB 1 MHQVDNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
 QY 61 PPAPATVAPPPPPAAANTPNAQPDNAPPPADPNAPPPPPVPIAPNAPQVRIQNVGGF 120
 DB 61 PPAPATVAPPPPPAAANTPNAQPDNAPPPADPNAPPPPPVPIAPNAPQVRIQNVGGF 120
 QY 121 SFALPAGWVESDAHFDYGSALLSKTTGDPFPQGPVPPVANDTRIVLGRDLQKLYASAEA 180
 DB 121 SFALPAGWVESDAHFDYGSALLSKTTGDPFPQGPVPPVANDTRIVLGRDLQKLYASAEA 180
 QY 181 TDSKAAARLGSDMGFEFYPYPTGTRINQTVSLDANGVSGSASYEVKFSKPGNQIWT 240
 DB 181 TDSKAAARLGSDMGFEFYPYPTGTRINQTVSLDANGVSGSASYEVKFSKPGNQIWT 240
 QY 241 GVIGSPAANAPDAGPPQRFVFWLGTANNPVYDKGAALAESIRPLVAPPAPAPAPAP 300
 DB 241 GVIGSPAANAPDAGPPQRFVFWLGTANNPVYDKGAALAESIRPLVAPPAPAPAPAP 300
 QY 301 APAPAPAGEVAPTPTTPTPQRTPLA 325
 DB 301 APAPAPAGEVAPTPTTPTPQRTPLA 325

RESULT 3
 ID APA MYCAV STANDARD; PRT; 381 AA.
 AC Q48919;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE Alanine and proline-rich secreted protein (45/47 kDa
 DE antigen) (Fibronectin attachment protein) (FAP-A).
 GN APA.
 OS Mycobacterium avium.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1764;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=101;

RX MEDLINE=97011577; PubMed=8858587;
 RA Schorey J.S., Holati M.A., Ratliff T.L., Allen P.M., Brown E.J.;
 RT "Characterization of the fibronectin-attachment protein of
 RT Mycobacterium avium reveals a fibronectin-binding motif conserved
 RT among mycobacteria";
 RL Mol. Microbiol. 21:321-329(1996).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U53585; AAB50543.1; -;
 KW Antigen; Signal.
 FT SIGNAL 1 32 POTENTIAL.
 FT CHAIN 33 381 ALANINE AND PROLINE-RICH SECRETED
 FT CHAIN 33 381 PROTEIN APA.
 SQ SEQUENCE 381 AA; 38210 MW; 0DACB9A22AA11D2B CRC64;
 Query Match 57.1%; Score 985; DB 1; Length 381;
 Best Local Similarity 55.2%; Pred. No. 2.7e-43;
 Matches 218; Conservative 27; Mismatches 66; Indels 84; Gaps 12;
 QY 1 MHQVDNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPP----- 55
 DB 1 MDQVEATSTRKGLWTLTATTVSGASAVVIALPETSHADPEVTPVPPSTATTTPHRRR 60
 QY 56 -----STAAPPAP-----TPVAPPPPPAAANTPNAQPDNAA-PPPA 93
 DB 61 RIGQPADNAQAGAPAPANGQORPPRRRMTTRAPPAGAPPNGAPPAENGAPPPV 120
 QY 94 DPNAPPPVIAAPNAPQVRIQNVGGFSPALPAGWVESDAHFDYGSALLSKTTGDPPPP 153
 DB 121 DPNAPPPPADPNAG--SIPN-----SYVLPAWVESDASHLDYGSALLSKVTGPPMP 172
 QY 154 GQPPVANDTRIVLGRDLQKLYASAEATDSKAAARLGSDMGFEFYPYPTGTRINQETVSLD 213
 DB 173 DQPPVANDTRIVMGVDQKLYASAEANNAKAAVGLGSDMGFEFYPYPTGTRINQDSTPLN 232
 QY 214 -ANGVSGSASYEVKFSKPGNQIWTGVIQSPAAANAPDAGPPQRFVFWLGTANNPV 272
 DB 233 GANGSTGSASYEVKFSKPGNQIWTGVIQSPAAANAPDAGPPQRFVFWLGTANNPV 287
 QY 273 KGAALAESIR---PLVAPPAP-----APAPAP- 300
 DB 288 KVAALAESIQAWTPPPAPPAAGGPGAPAGAPAGAPAGAPAGAPAGAPAGAPAG 347
 QY 301 -----APAP-----APAGEVAPTPTTPTPQRTPLA 325
 DB 348 APAPAGAPAGAPAGAPAGAPAGAPAGAPAGAPAGAPAGAPAGAPAGAPAG 381
 RESULT 4
 ID APA MYCLE STANDARD; PRT; 287 AA.
 AC P46842; O32905; Q9RSV6;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alanine and proline-rich secreted protein (45/47 kDa
 DE antigen) (Fibronectin attachment protein) (FAP-L).
 GN APA OR MODD OR ML2055 OR MLCB1788.01C.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94086110; PubMed=8262636;


```
RA Wiele B., van Agterveld M., Janson A., Clark-Curtiss J.E.,
RA Rinke de Wit T.F., Harboe M., Thole J.;
RT "Characterization of a Mycobacterium leprae antigen related to the
RT RT characterized Mycobacterium tuberculosis protein MP732.";
RL Infect. Immun. 62:252-258(1994).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=95310024; PubMed=7790081;
RA Schorey J.S., Li Q., McCourt D.W., Bong-Mastek M., Clark-Curtiss J.E.,
RA Ratliff T.L., Brown E.J.;
RT "A Mycobacterium leprae gene encoding a fibronectin binding protein is
RT used for efficient invasion of epithelial cells and Schwann cells.";
RL Infect. Immun. 63:2652-2657(1995).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN-TN;
RC MEDLINE=21128732; PubMed=11234002;
RX Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Henshaw K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy J., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whithead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
RN [2]
RN SUBCELLULAR LOCATION: Secreted.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76501; -; NOT ANNOTATED_CDS.
CC DR EMBL; AL583924; CAC31010.1; -.
CC DR EMBL; AL008609; CAA15433.1; -.
CC DR PIR; B87166; B87166.
CC DR Leproma; M22055; -.
CC KW Antigen; Repeat; Signal; Complete proteome.
CC FT SIGNAL 1 39
CC FT CHAIN 40 287
CC FT ALANINE AND PROLINE-RICH SECRETED PROTEIN
CC FT APA.
CC FT DOMAIN 76 96 3 X 4 AA REPEATS OF D-P-N-A.
CC FT REPEAT 76 99 1.
CC FT REPEAT 83 86 2.
CC FT REPEAT 93 86 3.
CC FT CONFLICT 33 34 LP -> FR (IN REF. 1).
CC FT SEQUENCE 287 AA; 29559 MW; 7FC860A8B3026CFB CRC64;
CC -----
Query Match 56.8%; Score 980; DB 1; Length 287;
Best Local Similarity 66.8%; Pred. No. 3.7e-43;
Matches 199; Conservative 21; Mismatches 64; Indels 14; Gaps 3;
QY 1 MHOVDNLTTRKGRLAALAIAMASASLVTVAVPATANADREPAPPVTTAASPSTAAA 60
DB 1 MNQVDLDSHRRGLWAILAIAVVASASFTMPLPAAANADPAFL-----PPSTATA 51
QY 61 PPAPATVPAPPVPPAANTNAQGGPNAPPPADNAPPVPPVIAENAPQVRINDPVGGF 120
DB 52 APSPAQEITLPLPGAPVSEAPQGDNA--PSLDENAPVPLAVDNDAG--RITNAVGGF 106
QY 121 SFALPAGWVESDAAHFDYGSALLSKTGDPPPGPPGVANDTRIVLGLDQKLYASAEA 180
DB 107 SFVLPAWGVSEASHLDYGSVLLSKAIEQPPVILGQFTVAVDTTRIVLGLDQKLYASAEA 166
QY 181 TTSKAAARLGSMDGFFMYPPTGRINQETVSDANGVSGSASYEVKFSDFSKNGQIWT 240
DB 167 DNKAARVLGSMDGFFLPPPTGRINQETIPLHANGIAGSASYEVKFSDFSKNGQICT 226
```

```
QY 241 GVIGSPAANADAGPPQRFVWVLGTANNPVDKGAALKAEISIRLVAPPAPAPAPA 298
DB 227 SVVGSPAASPTDPVGPQRWFWVVLGTNNPVDKGAALKAEISIRSEMAPIPASVSAPA 284
```

```
RESULT 5
GPI_CHLRE STANDARD; PRT; 555 AA.
AC Q9FP06; Q03927;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
DE glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=211159092; PubMed=11258910;
RA Farris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RN PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT encoding cell wall hydroxyproline-rich glycoproteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -1- FUNCTION: Major component of the outer cell wall M6 (crystalline)
CC layer.
CC -1- SUBUNIT: Associates with GP2 and GP3.
CC -1- PTM: N-glycosylated and O-glycosylated.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF309494; AAG45420.1; -.
CC DR EMBL; M58496; AAA69706.1; ALT_SEQ.
CC DR GlycoSuiteDB; Q9FP06; -.
CC DR InterPro; IPR003882; Pistil extensin.
CC DR PRINTS; PR01218; PSTLXTENSIN.
CC KW Glycoprotein; Repeat; Signal; POTENTIAL.
CC FT SIGNAL 1 29
CC FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
CC FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSX REPEATS.
CC FT DOMAIN 259 279 POLY-PRO.
CC FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT SEQUENCE 555 AA; 54219 MW; 6A584A9045502F5 CRC64;
CC -----
Query Match 14.1%; Score 243; DB 1; Length 555;
Best Local Similarity 24.6%; Pred. No. 6.8e-06;
Matches 70; Conservative 21; Mismatches 69; Indels 124; Gaps 8;
```

```
QY 39 ADPEAPPPVTTAASPPSTAAAPPATPV--APPPAAANTENAOPGDPNAPPADN 96
DB 203 APPSPAPPVPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 262
QY 97 APPPPVIAENAPQVRINDPVGGFSPALPAGWVESDAAHFDYGSALLSKTGDPPPG 156
```



```
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1."
RL J. Gen. Virol. 69:1531-1574(1998).
CC -!- FUNCTION: Tegument protein.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EH-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; X14112; CA32311.1; -.
DR PIR; I30085; WMBEH6.
DR InterPro; IPR006928; Herpes teg N.
DR Pfam; PF04843; Herpes teg N; 1.
DR Pfam; PF03586; Herpes_UL36; 1.
KW Repeat.
FT DOMAIN 2911 2980 35 X 2 AA TANDEM REPEATS OF P-Q.
SQ SEQUENCE 3164 AA; 335857 MW; CC5D31FF4F9FE3F4 CRC64;
Query Match 12.4%; Score 213.5; DB 1; Length 3164;
Best Local Similarity 26.5%; Pred. No. 0.001;
Matches 91; Conservative 17; Mismatches 117; Indels 119; Gaps 15;
QY 10 RKKGRLALATAMASA-----SLTVAVATATADPEPAPVPTTAAAP 61
DB 2683 RHRRARSARATQASATQGMRRPALDVTAVPTDFARP-PAPPKPEPAPHALVSGVP 2741
QY 62 -----PAPATPVAP-PPPAANTPNACQDPAAPPPADPNAPPPVIAAPAPQ 110
DB 2742 LPLGQAAGQASPALPIDVPPVATGT--VLPGGENRPLTSGPAPTFRVPGVGPQR 2799
QY 111 VRIDNPVGFSGFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIVLGR 170
DB 2800 -RLTRP-----AVASLSSESRESLSPWDP---ADPTAPVLGR- 2832
QY 171 DQLYASAEATDSKAARLGSMDGMEFYVPYPTGTRINQETVSLDANGVSGSASYEVKFS 230
DB 2833 -----NPAPTSSTSPAG----- 2844
QY 231 PSKPNQGIWTGVIGSPAANAPDAGPPQRFVYVWLGT-----NNPVDKGAALAESIR 284
DB 2845 PSPPPPAV-----QPVAPPSTSGPPPTLTLEGVAPGPGVSRRTTRQPVAPPTTSAR 2898
QY 285 P-----LVAPPAPAPAEAPAPAPAPAGEVAPTPTTPO 320
DB 2899 PRGHLTVSRSLAPQPPQPP-QPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 2940
RESULT 8
T2D3 HUMAN
ID T2D3 HUMAN STANDARD; PRT; 1083 AA.
AC 000268; Q99721; Q9BR40; Q9BX42;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcription initiation factor TFIID 135 kDa subunit (TAFII-135)
DE (TAFII135) (TAFII-130) (TAFII130).
GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```



```
FT SEQUENCE 802 AA; 85844 MW; 592B5975EE20F77F CRC64;
SQ SEQUENCE 802 AA; 85844 MW; 592B5975EE20F77F CRC64;

Query Match 11.8%; Score 205.5; DB 1; Length 802;
Best Local Similarity 22.9%; Pred. No. 0.00072;
Matches 78; Conservative 27; Mismatches 112; Indels 123; Gaps 11;

QY 33 VPATANADPPAPVPVPTAASPPSTAAA-----PPAPATPVAPPP----- 72
DB 343 VPRPLNKNRSPSSVNTPTSPQPPAKSCAWFTSNFSLPSPIMISSPPGKATGRPV 402
QY 73 PAAANTFNAQ-----PGDPNAA-----PPADNAPPPPVVIAPNAPQPVRI 118
DB 403 PUCVSSVPQMPSPPTAPNGSLDSVTYPVSPPTSGPAAPPPPPPPPPPPPPPP 454
QY 119 GFSFALPAGWVESDAAHFDVGSALLSKTTGDPDPGPGOP-PPVANDTIVLGR 177
DB 455 -----PPLPPPLPLPLAS----- 467
QY 178 AEATDSKAAARLGDGMGEFYMPYPTGTRI-----NQETVSLDANGVSGSAS 228
DB 468 -----LSHCSQASPPPTPLASTPSSKPSVLPSPSAGAPASAEETPLN 515
QY 229 SDPSKPNQIWTGIVGSPAANAPDAGPPQRFVWVLTANNPVVDKGAALAE 288
DB 516 SSASEP-----GLQASQPAESPTPOGLVLPAPPPPPPPPLPSGAYASAL 566
QY 289 ---PPAPAPAPAPAPAPAGVAPTPTTPTPTORTLPA 325
DB 567 PPPPPPLPSTGPPPPPPPPPPPLPNOAPPPPPPPPPPPPPPPPPPPPP 606

RESULT 11
TEGU_HSVBEB STANDARD; PRT; 3421 AA.
ID AC P28955;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein.
GN 24.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316 (1992).
CC -!- FUNCTION: Tegument protein.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M86664; AB02459.1; -.
DR F1R; G36797; W28EH6.
DR InterPro; IPR005928; Herpes_teg_N.
DR InterPro; IPR005210; Herpes_UL36.
DR Pfam; PF04843; Herpes_teg_N; 1.
DR Pfam; PF03586; Herpes_UL36; 1.
SQ SEQUENCE 3421 AA; 367078 MW; 5075EFE4739BB7AC CRC64;

Query Match 11.8%; Score 203; DB 1; Length 3421;
Best Local Similarity 23.9%; Pred. No. 0.0036;

Matches 96; Conservative 31; Mismatches 152; Indels 122; Gaps 15;
QY 6 PNLTRKGRLLAALAAAMASASLVTVAVPATANADPPAPVPVPTTAAASP-----STAAA 60
DB 2505 PTLPPKAAAPLPSPSDASAINSGPKVFKYTFGNKSAVPPSPVAPPTLPPAPPLPQSTSKAAS 2564
QY 61 PPAPATPVAPPPPPAAANTPNAQPGDPAAPPADPNAP-----PPPTLPPAPLPQSTSKAASGPPPTLPPAPP 108
DB 2565 GPPTLPPAPPLP--QSTSKAASG-----PPPTLPPAPLPQSTSKAASGPPPTLPPAPP 2617
QY 109 QVRINDNPVGGF-----SFALPAGWVESDAA----- 134
DB 2618 LPQSTSKAASGATSDSGKTLTLDVKTQSKDKVVPVPTDKPSTTTTFAALKQSDASKPP 2677
QY 135 -----HFDYGSALLSKTTGDDP-----PPG-OPP-----PVANDTIV 166
DB 2678 TTAIGHQKQLGTFVTPKSGDXPTDNASAPVGVSPVTPDGTGAKPPPKDAPPVDDTK-- 2735
QY 167 LGRLOQKLYASAEATDSKAAARLGDGMGEF-YMPYPTGTRIQTIVSLDANGVSGSASYE 225
DB 2736 ----QPVRKSLPSQVRGGRPVIRPSLGLPKFTGPPG-----YTI PVHGLPPSDSNVT 2783
QY 226 VKFSDPSKPNQIWTGIVGSPAANAPDAGPPQRFVWVLTANNPVVDKGAALAE 285
DB 2784 QSTKEPPK-----PAVETPAAPAK-----SAAAPAAAPAKSAAAPAAP 2823
QY 286 L-VAPPPAPAPAPAPAPAPAGVAPTPTTPTPTORTLPA 325
DB 2824 AKSAAAPAAAPAKSAAAPAAAPAKSAAAPAAAPAKDQTKSA 2864

RESULT 12
CPN_DROME STANDARD; PRT; 865 AA.
ID AC Q02910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Calphotin.
DE Calphotin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540 (1993).
CC -!- FUNCTION: Might function as a calcium-sequestering "sponge" to
CC regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
CC of Ca(2+) per mole of protein.
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; hypodense compartment.
CC -!- TISSUE SPECIFICITY: Soma and axons of photoreceptor cells of
CC compound eyes and ocelli.
CC -!- DEVELOPMENTAL STAGE: Expressed early in photoreceptor cell
CC development.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; L02111; AAA28405.1; --
 DR EMBL; L05080; AAA28420.1; --
 DR PIR; A47282; A47282.
 DR PIR; A47283; A47283.
 DR FlyBase; Fggn0010218; Cpn.
 DR GO; GO:0005509; F:calcium ion binding; IDA.
 KW Calcium-binding.
 FT CONFLICT 36 A -> AVAPAVVA (IN REF. 2).
 FT CONFLICT 43 I -> T (IN REF. 2).
 FT CONFLICT 64 I -> V (IN REF. 2).
 FT CONFLICT 76 T -> A (IN REF. 2).
 FT CONFLICT 100 P -> PP (IN REF. 2).
 FT CONFLICT 127 VQ -> AP (IN REF. 2).
 FT CONFLICT 154 I -> V (IN REF. 2).
 FT CONFLICT 160 S -> T (IN REF. 2).
 FT CONFLICT 160 S -> T (IN REF. 2).
 FT CONFLICT 534 A -> E (IN REF. 2).
 FT CONFLICT 699 I -> T (IN REF. 2).
 FT CONFLICT 703 V -> L (IN REF. 2).
 FT CONFLICT 721 D -> E (IN REF. 2).
 FT CONFLICT 721 D -> E (IN REF. 2).
 SQ SEQUENCE 865 AA; 84781 NW; 2110417E0B0E7CFE CRC64;
 Query Match 11.4%; Score 197.5; DB 1; Length 865;
 Best Local Similarity 26.6%; Pred. No. 0.002;
 Matches 105; Conservative 30; Mismatches 127; Indels 133; Gaps 18;
 QY 15 LAALAIAAMASGLVTVAVPAT-----ANADP--EPAPVPTTAASPTAAAPP--- 62
 Db 230 VSVVATKELAAEPPVAVPATETPVVAPAAASPHVSVAPAVETAVVAVPSASTEPVAA 289
 QY 63 -----APATEVAPP--AAANTPAQGDNAAPPDAPNAPP----- 99
 Db 290 ATLTTAPTALAPVVAESQVAANIVATP--PTAPEP--ETIAPPVVAETPEVASVAVA 346
 QY 100 -----PPVIAPNAPQVRIDNPVGGFSFALPAGVVED-----AAHFDY 138
 Db 347 ETTTPVVPVVAES--PAPVAVTP-----PATLAVTDVDTASAVPELPVPIAPSPV 399
 QY 139 GSALLSKTTGD-----PPFGQPP--VANDT-----RIVGLRQKLYASAEAT 181
 Db 400 FSA-VAETPVDLAPVLPVFAAEPPVAVVAEETPETPAPASAPVTIAALDIPEVAVPIAA 458
 QY 182 DSKAAARLGSDMGFYMFPYGPTRINQETVSLDANGVSGSASYEVKFSKNGQIWTG 241
 Db 459 PSDAPAEAPSAAPIVSTPTTASVPETTAPPA-----AVTEP-----IDVS 501
 QY 242 VIGSPAANAPAGPQRFVWVLTANNVDKGAALAESIRPLVAPPAPAPAPAPAP 301
 Db 502 VLSEAAIETPVA-----PPVEVTEVAAD-----VAPPEAAADLIIEFV 541
 QY 302 PAPA-----PAGEVAPTPTTPTORTLP 324
 Db 542 EPPAPIPDOLLQOTTSVPVAVAEASTSSPIPTSLP 576

RESULT 13

EXTN_SORBI STANDARD; PRT; 283 AA.
 AC P24152;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Extensin precursor (proline-rich glycoprotein).
 GN HRGP.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACAD clade; Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;

RN SEQUENCE FROM N.A.
 RP TISSUE=Leaf;
 RX MEDLINE=91370882; PubMed=1893107;
 RA Raz R., Cretin C., Puigdomenech P., Martinez-Izquierdo J.A.;
 RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum
 vulgare";
 RL Plant Mol. Biol. 16:365-367(1991).
 CC -!- FUNCTION: Structural component in primary cell wall.
 CC -!- FTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYCOSYLATED.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X56010; CA439485.1; --
 DR PIR; S13383; S13383. Pistill_extensin.
 DR InterPro; IPR003882; Pistill_extensin.
 DR PRINTS; PR01218; PSTLXTENSIN.
 DR Repeat; Cell wall; Glycoprotein; signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 283 EXTENSIN.
 SQ SEQUENCE 283 AA; 29593 MW; 8D7FCD0D8ED2D90 CRC64;
 Query Match 11.4%; Score 197; DB 1; Length 283;
 Best Local Similarity 24.0%; Pred. No. 0.00075;
 Matches 71; Conservative 17; Mismatches 108; Indels 100; Gaps 12;
 QY 34 PATANADPPAPVPTTAASPTAAAPAPATPVAPPPBAANTPNAQPG-DENVAAPP 92
 Db 71 PPTVTPSPKTPP-PATPKTPPTVTPSPKSPVYPPPPKASTPTVTPSPKPKATPP 129
 QY 93 ADPNAPPVPIAPNAPQVRIDNPVGGFSFALPAGVVEDAAHFDYGSALLSKITGDPFF 152
 Db 130 TYP-TPKPKATPP-PVYTPSP-----KPPVTKPT 160
 QY 153 PQQPPVANDTRIVGLRQKLYASAEATDSKAAARLGSDMGFYMFPYGPTRINQETVSL 212
 Db 161 PKPTPPV-----YTPNPKPPVTK----- 178
 QY 213 DANGVSGSASYEVKFSKNGQIWTGVIKSPAAANAPDAGPPQRFVWVLTANNPVD 272
 Db 179 -----PPTHVPSKPTSKTPPVYT-----PSKPKPKSP-----TYTTPKPKAT 221
 QY 273 KGAA-----KALASIRPLVAPP---PAPAPAPAPAPAPAPAGEVAPTPTTTP 319
 Db 222 KPPTSTPTPKPTPTPTVPPQAHPTTKPAPKSPAPPTP-PTYTTPVSGHTSPSP 276

RESULT 14

EXTN_TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).
 GN HRGPNT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=cv. Xanthi; TISSUE=Leaf;

MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 glycoprotein gene in lateral root initiation.";
 RL Genes Dev. 3:1639-1646(1989).
 CC -!- FUNCTION: Has a specialized structural function, possibly in
 CC the mechanical penetration of the cortex and epidermis of the
 CC main root.
 CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -!- FTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYCOSYLATED.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch)
 CC -----
 DR EMBL; X13885; CAA32090.1; -.
 DR FIR; S06733; S06733.1; -.
 DR InterPro; IPR000480; Glutelin.
 DR PRINTS; PR00211; GLUTELIN.
 DR Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 KW SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 620 EXTENSIN.
 FT REPEAT 70 73 H-A-P-P.
 FT REPEAT 148 151 H-A-P-P.
 FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 229 235 2.
 FT REPEAT 236 242 2.
 FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.
 FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.
 SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;
 Query Match 11.2%; Score 193; DB 1; Length 620;
 Best Local Similarity 25.0%; Pred. No. 0.0024;
 Matches 83; Conservative 21; Mismatches 102; Indels 126; Gaps 15;
 QY 30 TVAVPATANA-DPEP-----APVPTTAASPSTAAAPPAPA-----TPV-APPPPA 74
 Db 264 TYSPPPPAYAQSPQSPSYSPPPPSPISSPPPPSYSPPPPPSYSPPPPTTPTTSPPPPA 323
 QY 75 AANTNAQGDGNAAPPPADP-NAPPPVVIAP-----NAPQVEIDNPVGFS 121
 Db 324 YSPPTYSPPPTYPLFLRSSPISSPPPVYSPPPPSYSPPPPYLPSPSSPPPSFS 393
 QY 122 FALPAGWVESDAHFYDGSGALLSKTKTGDPFPF-----GQPPVANTRIVLGRLDQ 172
 Db 384 -PPPPTYEQSPPPPAYSPPLPAPTYSPPPTYSPPPTYSPPPTYPQQPPPL-----429
 QY 173 KLYASAEATDSKAARLGSMDGEFYMGEGRIHQETVSLDANGVSGSAYVEVKFSDPS 232
 Db 430 -----PPTYSPPPPAYSPPPPTYSPPPT-----YSGP- 458
 QY 233 KPNGQIWTGVIGSPAANPADGPQRWFVMGTANNPDVFXGAALAESIRPLVAPPPA 292
 Db 459 -----PPAYAQPPPPPTTY-----SP-----EPPA 478
 QY 293 PAPAPAEAPAPAPAGEVAFTTT---PTPQR 321
 Db 479 YGPPPPSPSIYSPPP-QVQLPFTPTSPPPRRR 509
 RESULT 15
 ID HCN4_RAT STANDARD; PRT; 1198 AA.
 AC QJUKA7; Q9QZW4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

KW Potassium channel; Potassium transport; Sodium transport;
KW CAMP; CAMP-binding; Transmembrane; Glycoprotein; Sodium channel.
FT DOMAIN 1 266
FT TRANSMEM 267 314
FT TRANSMEM 294 314
FT TRANSMEM 315 340
FT TRANSMEM 341 361
FT TRANSMEM 369 389
FT TRANSMEM 390 420
FT TRANSMEM 421 441
FT TRANSMEM 465 486
FT TRANSMEM 497 517
FT TRANSMEM 518 1198
FT TRANSMEM 209 260
FT NP_BIND 595 710
FT DOMAIN 799 1074
FT CARBOHYD 458 458
FT CONFLICT 404 404
SQ SEQUENCE 1198 AA; 128760 MW; 6B92B8F9452F760F CRC64;

Query Match 11.2%; Score 192.5; DB 1; Length 1198;
Best Local Similarity 23.8%; Pred. No. 0.0047;
Matches 98; Conservative 37; Mismatches 124; Indels 153; Gaps 19;

QY 10 RKKGFLAALAIAMASASLVT-----VAVPATA-----NADPEPAPPVPTTAASPP 55
DB 824 RHPRKLQSLIFALGSASPSPSQVDTFSSSFHIOQLAGFSAPGLSLPLPSSSSPP 883
QY 56 STA-AAPPAPA-----TPVAP-----70
DB 884 FGACSSPPAPTSTSTAATTGFGHEHKALGSLSSDSPLLTLPQCARSPQAAQPPPP 943
QY 71 -----PPPAANTNAQPG-----DPN-AAPPADENAPPP-----101
DB 944 LFGARGGLLEHFLPPPPSPSSSPSGQLGQPPGELSPGLAAGPSTPTETPRPERPS 1003
QY 102 --VIAPNAPQVRINDPNVGGFSFALPAGWVESDAHF-----DYGSAALSKTTGDP 150
DB 1004 FMAGASGASPVAF-TPRGGLS---PCHSPGPPTTPSPAPPASGSHGSLI-----P 1053
QY 151 PPGGPPPPVANDTR-----IVLGRLDQKL---YASAEATDSKAAARL-----GSDMGE 195
DB 1054 PASSPPPPQVQRRGTPTPTFGRLTQDLKLISQPALPQDGAQTLRRASPHSSGESMAA 1113
QY 196 FVMPYVPGTRINQETVSLDANGVGSASYYEVKFSQPKNGQIWTGVIGSPAANADAPG 255
DB 1114 FSL-YP-----RAGGSSSG-----GLGPPGPYPYGAIPGQHVTLPRKTISSGSLP 1157
QY 256 PQRFVFWLGTANNPVDKAAKAAEASIRPLVAPPAPAPAPAPAPAPAPA 307
DB 1158 P-----PLSLFGARAASSGGPPLTAAPQREPGARSEPRVKLPS 1196

Search completed: July 7, 2004, 18:21:51
Job time : 28.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 19:18:15 ; Search time 36.5 Seconds
(without alignments)
2809.409 Million cell updates/sec

Title: US-10-720-192-2

Perfect score: 1726

Sequence: 1 MHQVDPNLTTRKRLAALAI.....PAGEVAPTTPTTPTQRTLPA 325

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phage:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_virus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1539	89.2	286	2 Q9R640	Q9R640 mycobacteri
2	1053.5	61.0	368	2 Q9F4H9	Q9F4H9 mycobacteri
3	568	32.9	194	2 Q9V553	Q9V553 mycobacteri
4	256	14.8	3247	12 Q65553	Q65553 bovine herp
5	248.5	14.4	698	10 Q9ASK4	Q9ASK4 oryza sativ
6	248	14.4	3288	12 Q7T5D9	Q7T5D9 simian herp
7	244	14.1	511	6 Q95JD0	Q95JD0 sus scrofa
8	244	14.1	565	6 Q95JD1	Q95JD1 sus scrofa
9	241.5	14.0	745	16 Q89X06	Q89X06 bradyrhizob
10	240.5	13.9	3326	12 Q7T591	Q7T591 simian herp
11	238	13.8	1269	10 Q8W5K6	Q8W5K6 oryza sativ
12	238	13.8	1269	10 Q7XH56	Q7XH56 oryza sativ
13	237.5	13.8	1188	10 Q41805	Q41805 zea mays (m
14	231	13.4	668	5 Q8MXJ8	Q8MXJ8 caenorhabdi
15	230.5	13.4	584	16 Q82DU7	Q82DU7 streptomyc
16	228.5	13.2	464	10 Q41645	Q41645 volvox cart

17	227.5	13.2	652	4 Q8NAA1	Q8NAA1 homo sapien
18	227	13.2	598	16 Q8VKN7	Q8VKN7 mycobacteri
19	226.5	13.1	222	10 Q9S740	Q9S740 arabidopsis
20	226.5	13.1	609	16 Q8FM53	Q8FM53 corynebacte
21	225.5	13.1	420	5 Q9VZC2	Q9VZC2 corynebacte
22	225.5	13.1	652	10 Q94JZ6	Q94JZ6 arabidopsis
23	225	13.0	676	6 Q95JC9	Q95JC9 sus scrofa
24	224.5	13.0	3534	12 Q39266	Q39266 equine herp
25	223.5	12.9	652	10 Q91V48	Q91V48 arabidopsis
26	222	12.9	783	2 Q9XDH2	Q9XDH2 mycobacteri
27	222	12.9	3084	12 Q8UZ11	Q8UZ11 pseudorabie
28	218	12.6	409	10 Q9SBM1	Q9SBM1 volvox cart
29	217.5	12.6	839	16 Q9RX57	Q9RX57 deinococcus
30	217	12.6	784	13 Q90VB5	Q90VB5 gallus gall
31	217	12.6	801	5 Q23635	Q23635 caenorhabdi
32	216.5	12.5	496	16 Q7U8L8	Q7U8L8 synechococ
33	216.5	12.5	847	10 Q9XIB6	Q9XIB6 arabidopsis
34	215.5	12.5	817	3 Q07229	Q07229 saccharomyc
35	214.5	12.4	687	10 Q948Y7	Q948Y7 volvox cart
36	214.5	12.4	1016	10 Q9SPM0	Q9SPM0 zea mays (m
37	214	12.4	1204	3 Q8J0E5	Q8J0E5 varrowia li
38	214	12.4	1627	10 Q84ZL0	Q84ZL0 oryza sativ
39	213.5	12.4	760	10 Q9TOK5	Q9TOK5 arabidopsis
40	213.5	12.4	3122	12 P89459	P89459 herpes simp
41	212	12.3	753	5 Q93107	Q93107 acanthamoeb
42	211	12.2	889	16 Q9F2N5	Q9F2N5 streptomyc
43	209	12.1	1480	4 Q96Q04	Q96Q04 homo sapien
44	208.5	12.1	214	10 Q39763	Q39763 gossypium b
45	206.5	12.0	575	16 Q98F98	Q98F98 rhizobium l

ALIGNMENTS

RESULT 1

Q9R640 PRELIMINARY; PRT; 286 AA.
 ID Q9R640
 AC Q9R640
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE MPT-32=45 kDa CULTURE FILTRATE glycoprotein.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=96196153; PubMed=8626314;
 RA Dobos K.M., Khoo K.H., Swiderski K.M., Brennan P.J., Belisle J.T.;
 RT "Definition of the full extent of glycosylation of the 45-kilodalton
 glycoprotein of Mycobacterium tuberculosis.";
 RL J. Bacteriol. 178:2498-2506(1996).
 SQ SEQUENCE 286 AA; 28764 MW; 4867AD89FEB42493 CRC64;

Query March	89.2%	Score 1539;	DB 2;	Length 286;
Best Local Similarity	99.7%	Pred. No. 1.3e-89;		
Matches 285;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
QY	40	DPEAPPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTPNAQPGDPNNAAPPADPNAPP	99	
Db	1	DPEAPPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTPNAQPGDPNNAAPPADPNAPP	60	
QY	100	PVVIAPNAPQPVRIIDNPVGGFSGFALPAGWESDAAHFDYGSALLSKTTGDDPFPQPPPV	159	
Db	61	PVVIAPNAPQPVRIIDNPVGGFSGFALPAGWESDAAHFDYGSALLSKTTGDDPFPQPPPV	120	
QY	160	ANDTRIVLGRLOKLYASAEATDSKAARLGSMDGFEYMPYPTGTRINQINQVSLDANGVSG	219	
Db	121	ANDTRIVLGRLOKLYASAEATDSKAARLGSMDGFEYMPYPTGTRINQINQVSLDANGVSG	180	
QY	220	SASYTEVFPKSDSKNGQIWTGVIQSPNAPADAGPQRFWVWLGTTANNVVDKAAKAL	279	

Db 181 SASIYEVKFSKSPKNGQIWTGVTGSPANAPDAGPQRFVFWLGTANNPVDKGAAKAL 240
Q9F4H9 PRELIMINARY; PRT; 368 AA.
AC Q9F4H9
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Fibronectin-attachment protein FAP-P (Fibronectin-binding protein
DE P).
GN FAPP.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5781;
RA Secott T.E., Lin T.L., Wu C.C.;
RT "Fibronectin-attachment protein is necessary for efficient attachment
RT and invasion of epithelial cells by Mycobacterium avium subsp.
RT paratuberculosis.";
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF395912; AAL59141.1; -
FT NON_TER 1 1
FT NON_TER 194 194
SQ SEQUENCE 194 AA; 19339 MW; 3ABE7C8C93DF292 CRC64;
Query Match 32.9%; Score 568; DB 2; Length 194;
Best Local Similarity 63.9%; Pred. No. 1.3e-28;
Matches 124; Conservative 14; Mismatches 16; Indels 40; Gaps 7;
QY 162 DTRIVLGRDLQKLYASAEATDSKAAARLGSMDGFEYMPYPTGTRINQETVSLD-ANGVSGS 220
Db 1 DTRIVLGRDLQKLYASAEANNAKAAVRLGSDMGFEYMPYPTGTRINQDSTPLANGANGSTGS 60
QY 221 ASYIEVKFSDPSKPNQIWTGVTGSPANAPDAGPQRFVFWLGTANNPVDKGAAKALA 280
Db 61 ASYIEVKFSDPSKPNQIWTGVTGSPANAPDAGPQRFVFWLGTANNPVDKGAAKALA 115
QY 281 ESIR---PLVAPPAP-----APAPAP-----APAP--APA 307
Db 116 ESIAQWTPPAAPPAAGPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 175
QY 308 GEVAPTPTPTPQOR 321
Db 176 VEVSPTP-TPTPQ 188
RESULT 4
Q65553 PRELIMINARY; PRT; 3247 AA.
AC Q65553
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE UL36.
DE UL36.
GN UL36.
OS Bovine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
RA Thiry E., Paces V.;
RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=COOPER;
RA Schwyzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA LaBoissiere S., Misra V., Vlcek C., Paces V.;
RL Vet. Microbiol. 0:0-0(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;

Q9F4H9 PRELIMINARY; PRT; 194 AA.
AC Q9F4H9
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Fibronectin-attachment mutant protein (Fragment).
GN FAPP.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5781;
RA Secott T.E., Lin T.L., Wu C.C.;
RT "Fibronectin-attachment protein is necessary for efficient attachment
RT and invasion of epithelial cells by Mycobacterium avium subsp.
RT paratuberculosis.";
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF395912; AAL59141.1; -
FT NON_TER 1 1
FT NON_TER 194 194
SQ SEQUENCE 194 AA; 19339 MW; 3ABE7C8C93DF292 CRC64;
Query Match 32.9%; Score 568; DB 2; Length 194;
Best Local Similarity 63.9%; Pred. No. 1.3e-28;
Matches 124; Conservative 14; Mismatches 16; Indels 40; Gaps 7;
QY 162 DTRIVLGRDLQKLYASAEATDSKAAARLGSMDGFEYMPYPTGTRINQETVSLD-ANGVSGS 220
Db 1 DTRIVLGRDLQKLYASAEANNAKAAVRLGSDMGFEYMPYPTGTRINQDSTPLANGANGSTGS 60
QY 221 ASYIEVKFSDPSKPNQIWTGVTGSPANAPDAGPQRFVFWLGTANNPVDKGAAKALA 280
Db 61 ASYIEVKFSDPSKPNQIWTGVTGSPANAPDAGPQRFVFWLGTANNPVDKGAAKALA 115
QY 281 ESIR---PLVAPPAP-----APAPAP-----APAP--APA 307
Db 116 ESIAQWTPPAAPPAAGPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 175
QY 308 GEVAPTPTPTPQOR 321
Db 176 VEVSPTP-TPTPQ 188
RESULT 4
Q65553 PRELIMINARY; PRT; 3247 AA.
AC Q65553
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE UL36.
DE UL36.
GN UL36.
OS Bovine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
RA Thiry E., Paces V.;
RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=COOPER;
RA Schwyzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA LaBoissiere S., Misra V., Vlcek C., Paces V.;
RL Vet. Microbiol. 0:0-0(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;

```

RA Schwytzer M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=JURA;
RX MEDLINE=97164286; PubMed=9010999;
RA Schwytzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA LaBoissiere S., Misra V., Vlcek C., Paces V.;
RT "Gene contents in a 31-kb segment at the left genome end of bovine
RT herpesvirus-1.";
RL Vet. Microbiol. 53:67-77(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwytzer M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwytzer M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 278205; CAB01605.1; -.
DR EMBL; AJ004801; CA06097.1; -.
DR InterPro; IPR006928; Herpes teg N.
DR InterPro; IPR005210; Herpes_UL36.
DR Pfam; PF04843; Herpes_teg_N; 1.
DR Pfam; PF03586; Herpes_UL36; 1.
SQ SEQUENCE 3247 AA; 332190 MW; 3AEAA72F8F001F6A CRC64;

Query Match 14.8%; Score 256; DB 12; Length 3247;
Best Local Similarity 31.2%; Pred. No. 1.4e-07;
Matches 105; Conservative 20; Mismatches 153; Indels 58; Gaps 15;

QY 8 LTR-----RKGLAALAIAMASASITVAVPATANADPEAPVPTTAAASPPSTAAAPPA 63
DB 2584 LTRRTGPKSPFALPMWRPPPSASLSAAFPVPPGRLPPAPLPPLPPAPLPPLPP 2643

QY 64 PATVAPP-PPAANTENAGQCDNAPPPADNAP-----PPVIANAPQPVRIID 114
DB 2644 PAPLPPLPPPLPPPLPPAPLPSTAPVAPPLPPPLPPALTALTPATPTAPLPPLP 2703

QY 115 NPVGFSFALPAGWVESDAHFYGSALLSTTGDPPFPFGQP-PPVANDTRIVLGRDQK 173
DB 2704 API---TVLVAP-VFAPAP-----IPAPATPAPATPAPLPPLPPPDGNGALSATRR 2754

QY 174 LYASAEATDSKAAA-----RLGSDMGEFYMPGTRINQETVSLDANGVSGSASYEVKFS 229
DB 2755 PTRRAGARKSLPAAQPRQLRLRSRSPASVPAGS-----ELVPPSPGGALGSPPSFV---- 2806

QY 230 DPSPKNGQIWTGVGSPAANAPDAGPPQRFVFWLGTANKPVDKGAALAESIRPLVAP 289
DB 2807 -PSRPPSL-----EPVGLP-LPPSR-----VQAPVDAPAPPAPERAPPAP 2848

QY 290 P-PAPAPAPAPAPAPAGVAPTPTPTPTPTPT 324
DB 2849 ERPAPPAPERAPPAPERAPP-----PAPERAP 2880

RESULT 5
Q9ASK4 PRELIMINARY; PRT; 698 AA.
ID Q9ASK4
AC Q9ASK4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN P0439B06.8.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone:PO439B06.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AP002882; BAB39873.1; -.
DR Gramene; Q9ASK4; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR008271; Ser_Thr_kin_AS.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR Hypothetical protein; ATP-binding; Kinase.
KW Serine/threonine-protein kinase; Transferase.
KW SEQUENCE 698 AA; 72733 MW; F386E46DCA9DAE6 CRC64;

Query Match 14.4%; Score 248.5; DB 10; Length 698;
Best Local Similarity 30.7%; Pred. No. 7.5e-08;
Matches 89; Conservative 20; Mismatches 134; Indels 47; Gaps 9;

QY 32 AVPATANADPEAPVPTTAAASPPSTAAAPAPATPVAPPPAAAANTPNACGCPNAAPP 91
DB 12 ATPATAAAPTATPTP---ATQPNATPADPSITPPAASPLFSAATPPPPQDPPSPSP 68

QY 92 PADPNAPPPVIAPNAPQPVRIIDNPVGFSFALPAGWVESDAHFYGSALLSKTTGDPP 151
DB 69 PAPVAVPPATVPVPPVPVASEPPT-----PSATLPPPPSPASVPVPTPTPKP 121

QY 152 FPCQPPVANDTRIVLGRDQKLYASAEATDSKAAARLGSMDGFEYV-PYPGTRINQETV 210
DB 122 SPVQQPEVA-----ASPPSPADLPPNPPAR--SDTPPVVQSPPPHRRSPRT 170

QY 211 SLDANGVSGSASYEVKFSKFNQIWTGVGSPAANAPDAGPPQRFVFWLGTANNP 270
DB 171 WAPPMAPSFGS-----PTKP-----SPASFSPIAGDP-----IITPNNP 206

QY 271 VDKGAALAESIRPLVAPPAPAPAPAPAPAGVAPTPTPTPTPT 320
DB 207 SSPLATPSAGSGTPVVT-PSAPVSGPPSGTAPATAADRSNKSLSPTQ 255

RESULT 6
Q7TSD9 PRELIMINARY; PRT; 3288 AA.
ID Q7TSD9
AC Q7TSD9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Very large tegument protein.
GN UL36.
OS Simian herpes B virus (Cercopithecoid herpesvirus 1) (Shbv).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10325;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E2490;
RX MEDLINE=22628476; PubMed=12743273;
RA Perelygina L., Zhu L., Zurkhuun H., Mills R., Borodovsky M.,
RA Hilliard J.K.;
RT "Complete Sequence and Comparative Analysis of the Genome of Herpes B
RT Virus (Cercopithecine Herpesvirus 1) from a Rhesus Monkey.";
RL J. Virol. 77:6167-6177(2003).
DR EMBL; AF533768; AAP41454.1; -.
SQ SEQUENCE 3288 AA; 342496 MW; 79C391EE8E6F7983 CRC64;

```



```
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RA MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005936; BAC45786.1; -.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR006665; OmpA/MotB.
DR InterPro; IPR003882; PstII_extensin.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01218; PSTIEXTENSIN.
DR ProDom; PD000001; Prot_kinase; 1.
KW Complete proteome.
SQ SEQUENCE 745 AA; 74545 MW; 155EDFC74DBCF6D CRC64;

Query Match 14.0%; Score 241.5; DB 16; Length 745;
Best Local Similarity 29.6%; Pred. No. 2.2e-07;
Matches 112; Conservative 16; Mismatches 136; Indels 115; Gaps 21;

Qy 15 LAALAIAM-----ASASIVTVAVPATANADPEP-----APP----- 46
Db 9 LATTALTMQLANTASHAQASPLVVAQAQOETGPDQKFKQPPKPGGAPPAAPARPA 68
Qy 47 VETTAAGPPSTAAPP---APATVAPPPAAAANTPNAQGDENAPPPADPNAP-PPPV 102
Db 69 APPAAAPPPHPPAAPPAAAPPPAAPP---PPAARPAAPPPPPPPPAAPPKQSPPPA 126
Qy 103 IAP--NAPQVRIQNVGSGFSAIPAGWVESDAHFYGSALLSKTTGDP-PPGQPPPV 159
Db 127 AAPQOCHAPT---PPPAAPPAARPAFTPPAPPAAQOHPPPPPPPAARPTPTPPPPPA 183
Qy 160 ANDTRIVLGRDQKLYASAEATDSKAA-----ARLGSOMGEYMPYVG-----TRI 205
Db 184 GPAARPT-----PAPTATPTFPVAPPPAAPTARPGS-----PAPAATPAPTTPA 227
Qy 206 NOETVSLDANGVSGSASYEVKFSKPNQIWTGIVGSPAANAPDAGPQRFVWVLG 265
Db 228 PTATPAPTATPAGSTP-----GAPPAGRCGAPPPG---VRPG 262
Qy 266 T---ANRPVDKGAALAEIRP--LVAPP-----PA--PAPAPAEAP--APAPAGEVA 311
Db 263 SPPAGSGPPAGCATPAPTTPAPGCTATPPSGRCGPASTPAPGAATPAPTATPAGGALT 322
Qy 312 PTP-----TPTPTORTLP 324
Db 323 PPPGRPGAGTTPGQGGTP 341

RESULT 10
ID Q7T591 PRELIMINARY; PRT; 3326 AA.
AC Q7T591;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE large tegument protein.
GN UL36.
OS Simian herpes B virus (Cercopithecine herpesvirus 1) (shbv).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_taxID=10325;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22607624; PubMed=12721804;
RA Ohsawa K., Black D.H., Sato H., Rogers K., Eberle R.;
RT "Sequence and genetic arrangement of the UL region of the monkey B
RT virus (Cercopithecine herpesvirus 1) genome and comparison with the UL
RL region of other primate herpesviruses."
RL Arch. Virol. 148:989-997(2003).
DR EMBL; AB09160; BAC58076.2; -.
SQ SEQUENCE 3326 AA; 345566 MW; 6B53E3860F43CDF0 CRC64;

Query Match 13.9%; Score 240.5; DB 12; Length 3326;
Best Local Similarity 29.1%; Pred. No. 1.3e-06;
Matches 93; Conservative 24; Mismatches 162; Indels 41; Gaps 9;

Qy 17 ALAIAAMASASIVTVAVPATANADPEPAPPTTTAASPESTAAAPAPATPVAP--PPPA 74
Db 2875 AAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAA 2934
Qy 75 AANTNAQCGDNAAAPPPADPNAPPPV--IAPNAPQVRIQNVGSGFSAIPAGWVES 132
Db 2935 AVPAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAAAPAA 2994
Qy 133 AAHFYGSALLSKTTGDPFPFGOP-----PPVANDTRIVLG-----RLDQKLYASAEATD 182
Db 2995 APAATPASPVPPTSLPTPPSKPPAFFQPSLATGGSVAPGGDFERRAPSRPTAAVPAAP 3054
Qy 183 SKAAAR-----LGSMDGEYMPYGPCTRIQNETVSLDANGVSGSASYEVKFSKPNQ 236
Db 3055 SRPPARRLARPAVSRSTESFALP-FDELARPTPEAPAPTETETAPVAERPAPEPPQ 3113
Qy 237 QIWTGIVGSPAANAPDAGPQRFVWVLGTTANNPVDKGAALAEIRPLVAPP---PA 292
Db 3114 R-----PPSPAAPDAGP-----AAASGSGGVPAPLGALVPGVAVRRQIDP 3157
Qy 293 PAPAPAEAPAPAPAGEVAP 312
Db 3158 PAPPREIPAPSPPPRSHAP 3177

RESULT 11
ID Q8W5K6 PRELIMINARY; PRT; 1269 AA.
AC Q8W5K6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OSJNBA0079B05.10 OR OSUNAA0079B05.2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_taxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Sasaki C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
RA Rambo T., Sasaki C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079179; AAL31655.1; -.
DR EMBL; AC116601; AAM08709.1; -.
DR Gramene; Q8W5K6; -.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR008973; C2 CALB.
DR InterPro; IPR003104; FH2.
DR Pfam; PF02181; FH2; 1.
DR SMART; SM00496; FH2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1269 AA; 138432 MW; 262B546481B25CA6 CRC64;
```

```

Query Match      13.8%; Score 238; DB 10; Length 1269;
Best Local Similarity 27.2%; Pred. No. 6.6e-07;
Matches 82; Conservative 20; Mismatches 121; Indels 78; Gaps 10;

QY 36 TANADPEAPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTNAOQDPNAAPP----- 91
DB 540 TAAAPPPPPPPPPPSGNKP--AFSPPPPPPPPPPPPLPQSNYASSQPPPPPPPPPPPPNC 597
QY 92 --PADPNAPPPVIAPN--APQVRIDNPVGGFSGFALPAGWVESDAAHFDYGSALLSKTT 147
DB 598 LVSPSPPPPPPPPPILPNRSVPPPPPPPPPPPLNHSVL----- 633
QY 148 GPPPPGPPPPVANDTRIVLGLDQKLYASAEATDSKAAARLGSMDGCFYMPYPTGTRINQ 207
DB 634 --PPPPPPPPPPPSLNLV-----PPPPAPGIGNKFPAPPPPPPPPPSSS 676
QY 208 ETVSIDANGVSGSASYEVKFSDFSKPNGQIWTGVIGSPAANAPDAGPQRFVWLGT 267
DB 677 RPTGTAATSSKGGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 724
QY 268 NNPDVKGAKALAEISIRPLVAPPAPAPA-----PAEPAPAPAPAGEVAPTPTTPTPQRTL 323
DB 725 NGP-----SAPAPPL--PPPLPAAANKRNPAPPPPPPLMTGKKAPAPAPPPPPPPQAPK 773
QY 324 P 324
DB 774 P 774

RESULT 12
QYXH56 PRELIMINARY; PRT; 1269 AA.
AC QYXH56;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DE Hypothetical protein.
GN OSUNBA0079305.10.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RT Chromosome 10.";
RL Science 300:1566-1569 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017051; AAP51898.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1269 AA; 138431 MW; 262E546481B25CA6 CRC64;

Query Match      13.8%; Score 238; DB 10; Length 1269;
Best Local Similarity 27.2%; Pred. No. 6.6e-07;
Matches 82; Conservative 20; Mismatches 121; Indels 78; Gaps 10;

QY 36 TANADPEAPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTNAOQDPNAAPP----- 91
DB 540 TAAAPPPPPPPPPPSGNKP--AFSPPPPPPPPPPPPLPQSNYASSQPPPPPPPPPPPPNC 597
QY 92 --PADPNAPPPVIAPN--APQVRIDNPVGGFSGFALPAGWVESDAAHFDYGSALLSKTT 147
DB 598 LVSPSPPPPPPPPPILPNRSVPPPPPPPPPPPLNHSVL----- 633
QY 148 GPPPPGPPPPVANDTRIVLGLDQKLYASAEATDSKAAARLGSMDGCFYMPYPTGTRINQ 207

Query Match      13.8%; Score 238; DB 10; Length 1269;
Best Local Similarity 27.2%; Pred. No. 6.6e-07;
Matches 82; Conservative 20; Mismatches 121; Indels 78; Gaps 10;

QY 36 TANADPEAPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTNAOQDPNAAPP----- 91
DB 540 TAAAPPPPPPPPPPSGNKP--AFSPPPPPPPPPPPPLPQSNYASSQPPPPPPPPPPPPNC 597
QY 92 --PADPNAPPPVIAPN--APQVRIDNPVGGFSGFALPAGWVESDAAHFDYGSALLSKTT 147
DB 598 LVSPSPPPPPPPPPILPNRSVPPPPPPPPPPPLNHSVL----- 633
QY 148 GPPPPGPPPPVANDTRIVLGLDQKLYASAEATDSKAAARLGSMDGCFYMPYPTGTRINQ 207

```

```

DB 634 --PPPPPPPPPSLNLV-----PPPPAPGIGNKFPAPPPPPPPPPSSS 676
QY 208 ETVSIDANGVSGSASYEVKFSDFSKPNGQIWTGVIGSPAANAPDAGPQRFVWLGT 267
DB 677 RPTGTAATSSKGGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 724
QY 268 NNPDVKGAKALAEISIRPLVAPPAPAPA-----PAEPAPAPAPAGEVAPTPTTPTPQRTL 323
DB 725 NGP-----SAPAPPL--PPPLPAAANKRNPAPPPPPPLMTGKKAPAPAPPPPPPPQAPK 773
QY 324 P 324
DB 774 P 774

RESULT 13
QYI805 PRELIMINARY; PRT; 1188 AA.
AC QYI805;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Extensin-like protein precursor.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B73; TISSUE=Pollen;
RA Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;
RT "Pex genes: pollen-specific genes with extensin-like domains.";
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z34465; CAA84230.1; -.
DR FIR; S49915; S49915.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007090; LRR_plant.
DR Pfam; PF00560; LRR; 3.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
SQ SEQUENCE 1188 AA; 120981 MW; 2C77C7F8D7130149 CRC64;

Query Match      13.8%; Score 237.5; DB 10; Length 1188;
Best Local Similarity 28.9%; Pred. No. 6.6e-07;
Matches 87; Conservative 30; Mismatches 123; Indels 61; Gaps 13;

QY 34 PATANADPEP--APPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTNAOQDPNAAPP 91
DB 569 PAPVASPPPPVKSPPPPTLVASFPFPPVKSPPPPA--PVASPPPPVKSP--PPTVASFP 624
QY 92 PADPNA--PPPVIANAPQVRIDNPVGGFSGFALPAGWVESDAAHFDYGSALLSKTTGDP 150
DB 625 PPAPVASSPPPMKSP--PPPTPVSSPPPPPEKSPPPPPPAKSTPPPEEYPTPTTSVKSSPP 682
QY 151 PFPQPPPVANDTRIVLGRDLQKLYASAEATDSKAAARLGSMDGCFYMPYPTGTRINQITV 210
DB 683 PEKSLDPP-----TLIPSPPPQEK--PTPSTPSKPPS-----SPEKP 718
QY 211 SLIDANGVSGSASYEVKFSDFSKPNGQIWTGVIGSPAANAPDAGPQRFVWLGTANNP 270
DB 719 SPPEKPVSSPPQ--TPKSSPPAPVSSPPPTPVSSPPPALAPVSSPP-----SVKSSP 768
QY 271 VDKGAALAEISIRPLVAPPAP-----APAPAE---PAPAPAPAGEVAPTPTTPTPQRTL 323
DB 769 PP-----APLSSPPPPAPQVKSPPPPVQVSSPPPPAPKSPPLAPVSSPPQVEKTS 817
QY 324 P 324
DB 818 P 818

```

```
RESULT 14
Q82DU7
ID Q8MXU8 PRELIMINARY; PRT; 668 AA.
AC Q8MXU8
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein K08D12.6.
GN K08D12.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Wu X., Gattung S.;
RA "The sequence of C. elegans cosmid K08D12.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Waterston R.;
RA Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC006672; AA098004.1; -
DR WormPep; K08D12.6; CE21047.
KW Hypothetical protein.
SQ SEQUENCE 668 AA; 62778 MW; FB9ACC5739DC398A CRC64;

Query Match 13.4%; Score 231; DB 5; Length 668;
Best Local Similarity 26.4%; Pred. No. 9e-07;
Matches 107; Conservative 23; Mismatches 137; Indels 138; Gaps 17;

QY 16 AALAIAAMASASLVTVAVPATA-----NADPEPAPPVPTTAAS----- 53
DB 169 AAAPAPADAPVEQAPVAVPAPAPAAAPDCGSAAPAPAPAPAPAPAPATDSGYRSKENSXGDE 228
QY 54 --PSTTAAPAPAP-----TPVAPPPPPAAANTNAQFGD----- 85
DB 229 QVTPAPAAAAPAPADAPVEQAPVAVPAPAPVPAAPDVECGSAAPAPAAAAPAPATDSGYRSK 288
QY 86 -----PNAAPPADPNAPPPPVIAAPQPVRIAPVGVGFSFALPA----- 126
DB 289 RNSYGDQVTPAPAAAAPADAPVEQAPVAVP-APAPAAAPDCGSAAPAPAAAAPATD 347
QY 127 -GWVESDAAHFDY----GSALLSKTTGPPPPGPPPPVANDTRIVLGRDOKLYASAEAT 181
DB 348 SGYRSKRNAYGDEQVTPAPAAAAPADAPV--EQAPVA-----VPAPAPTAAPAP 396
QY 182 DSKAARLGSDMG-----EFYMPYPCTR-----INQETVSLDANGVSGSASY 223
DB 397 DCGSAAPATDSGYRSKRNAYGDSQVTPAPAAAAPADAPVEQAPVAVPAPAPAAA--- 453
QY 224 YEYKFSDFSKNGQIWTGVIQSGAANAPDAGPPQPPQVFWLGTANNPVDKGAALAEIRPLVAPPPAPAPAPAP 280
DB 454 -----PAPDCGS-----AAPAAAAPAAAAPAA-----ATDSGYRSKRNAYG 488
QY 281 -ESIRPLVAPPAPAPAPAPAPAP-APAPAGEVAPTPTTTPQRT 322
DB 489 DEQVTPAPAAAAPADAPVEQAPVAVPAPAPAAAAPAPAPAPAPAT 533
```

```
RESULT 15
Q82DU7
ID Q82DU7 PRELIMINARY; PRT; 584 AA.
```

```
Q82DU7;
01-JUN-2003 (TrEMBLrel. 24, Created)
01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV4871.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005040; BAC72583.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 584 AA; 59261 MW; 7AA08D5F87EE2378 CRC64;

Query Match 13.4%; Score 230.5; DB 16; Length 584;
Best Local Similarity 29.6%; Pred. No. 8.4e-07;
Matches 94; Conservative 16; Mismatches 113; Indels 95; Gaps 16;

QY 29 VIVAVPA-----TANADPEPAPPVPTTAASPPSTAAAPAPATPVAPP--PPAAAN 77
DB 110 VRVGTTPAGTGAAGGGAAPPPPPPPP-----SPPRFVVEPPVQRPVAGVGPVPGD 165
QY 78 TPNACQGDPN-----AAPPADPNAPPPPPVIAAP-NAPOFVRIDNPVGVGFSFALPAG 127
DB 166 AQCGPPPRDKATHLLRAVAEFTPAQWTPQOSVAPAPAPAPAPVERPASTPTHVLA- 224
QY 128 WVESDAAHFDYGSALLSKTTGPPPPGPPPPVANDTRIVLGRDOKLYASAEATSKA-- 185
DB 225 LPESAATH-----QPEQVQAPAPKPKVPEG-----AEAVASEAEE 261
QY 186 -AARLGSDMBEFVMPYPGTRINQETVSLDANGVSGSASYEVKFSDPKPNQIWTGVIG 244
DB 262 TPVAGGTVSTFVLRVAVPEQAFWIV---APGLAQL-----FGAPTQ-----G 303
QY 245 SPAANAPDAGPPQPPVFWLGTANNPVDKGAALAEIRPLVAPPPAPAPAPAPAP 304
DB 304 APAPAGTVSPPTGAF-----GPA-----PTTATPP-PGPAPWEPAPAP 341
QY 305 APAGEVAPTPT---TPTP 319
DB 342 EPAPAPAPAPAOGRPTTP 359

Search completed: July 7, 2004, 18:23:02
Job time : 39.5 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:21:10 ; Search time 16.5 Seconds
(without alignments)
1016.875 Million cell updates/sec

Title: US-10-720-192-2
Perfect score: 1726
Sequence: 1 MHQVDNLTTRKGRLAALAI.....PAGEVAPTPTPTPTQRTLPA 325

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1726	100.0	325	1	US-08-382-184-2
2	1726	100.0	325	2	US-08-641-356-2
3	1726	100.0	325	3	US-09-132-528-2
4	1726	100.0	325	3	US-09-132-528-3
5	1726	100.0	325	3	US-08-875-494-2
6	1726	100.0	325	4	US-08-599-366-2
7	1726	100.0	325	4	US-08-599-366-3
8	1726	100.0	325	4	US-08-875-494-2
9	1726	100.0	325	4	US-08-510-031A-6
10	1726	100.0	325	4	US-09-985-372-2
11	1726	100.0	325	4	US-09-985-372-3
12	1726	100.0	325	4	US-08-818-112-53
13	1726	100.0	325	4	US-08-818-111-53
14	1726	100.0	325	4	US-09-086-556-53
15	1726	100.0	325	4	US-09-072-596-53
16	1726	100.0	325	4	US-09-072-967-53
17	1720	99.7	325	4	US-09-510-031A-2
18	1542	89.3	286	1	US-08-382-184-3
19	1542	89.3	286	2	US-08-641-356-3
20	1542	89.3	286	3	US-08-132-528-4
21	1542	89.3	286	3	US-08-875-494-3
22	1542	89.3	286	4	US-09-599-366-4
23	1542	89.3	286	4	US-08-875-494-3
24	1542	89.3	286	4	US-09-985-372-4
25	1539	89.2	285	4	US-09-510-031A-7
26	1499	86.8	652	4	US-09-072-596-350
27	1499	86.8	652	4	US-09-072-967-355

28	1499	86.8	802	4	US-09-056-556-214	Sequence 214, App
29	1499	86.8	802	4	US-09-072-596-209	Sequence 209, App
30	1499	86.8	802	4	US-09-072-596-346	Sequence 346, App
31	1499	86.8	802	4	US-09-072-967-214	Sequence 214, App
32	1499	86.8	802	4	US-09-072-967-351	Sequence 351, App
33	1499	86.8	802	4	US-09-287-849-10	Sequence 10, Appl
34	985	57.1	381	4	US-09-510-031A-5	Sequence 5, Appl
35	980	56.8	287	4	US-08-311-731A-96	Sequence 96, Appl
36	980	56.8	287	4	US-09-510-031A-4	Sequence 4, Appl
37	969	56.1	287	4	US-09-510-031A-8	Sequence 8, Appl
38	233	13.5	8991	4	US-08-714-741-32	Sequence 32, Appl
39	200.5	11.6	214	1	US-08-217-327-4	Sequence 4, Appl
40	200	11.6	408	1	US-07-609-716-65	Sequence 65, Appl
41	200	11.6	408	3	US-08-475-411A-65	Sequence 65, Appl
42	200	11.6	408	3	US-08-478-029A-65	Sequence 65, Appl
43	197.5	11.4	330	1	US-08-642-255-32	Sequence 32, Appl
44	197	11.4	538	4	US-09-616-288-43	Sequence 43, Appl
45	196.5	11.4	666	4	US-09-050-739-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1
US-08-382-184-2
; Sequence 2, Application US/08382184
; Patent No. 5714593
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MICROBACTERIAL PROTEINS.
; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINES
; TITLE OF INVENTION: AND FOR THE DETECTION OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER &
; ADDRESS: 1755 S. Jefferson Davis Highway, Suite 400
; City: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; FILING DATE: 01-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 22640720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-382-184-2

Query Match 100.0%; Score 1726; DB 1; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPVPVPTTAAASPSTAAA 60
DB 1 MHQVDNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPVPVPTTAAASPSTAAA 60

```
QY 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
Db 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
QY 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
QY 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
Db 301 APAPAPAGEVAPTPTPTPTQRTPLA 325

RESULT 2
US-08-641-356-2
; Sequence 2, Application US/08641356
; Patent No. 5866130
; GENERAL INFORMATION:
; APPLICANT: MICROBACTERIAL PROTEINS,
; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINES
; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT, P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/641,356
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/382,184
; FILING DATE: 01-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 22640720
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-641-356-2

Query Match 100.0%; Score 1726; DB 2; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDNLTTRKGRGLAALAIAAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
Db 1 MHQVDNLTTRKGRGLAALAIAAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
QY 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
Db 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
QY 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
QY 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
Db 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
```

```
QY 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
Db 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
QY 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
QY 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
Db 301 APAPAPAGEVAPTPTPTPTQRTPLA 325

RESULT 3
US-09-132-528-2
; Sequence 2, Application US/09132528A
; Patent No. 6221353
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/132,528A
; CURRENT FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: 08/641,356
; EARLIER FILING DATE: 1996-04-30
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-132-528-2

Query Match 100.0%; Score 1726; DB 3; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDNLTTRKGRGLAALAIAAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
Db 1 MHQVDNLTTRKGRGLAALAIAAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
QY 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPPAAAANTPNAOQDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
Db 181 TDSKAAARLGSDMGSEFYMPYPTGTRINQETVSLDANGVSGSASYEYVKSDFSKPQNGIWT 240
QY 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPQRFVFWLGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAE 300
QY 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
Db 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
```

```
Db 301 APAPAGEVAPTPTTPTQRTLPA 325

RESULT 4
US-09-132-528-3
; Sequence 3, Application US/09132528A
; Patent No. 6221353
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/132.528A
; CURRENT FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: 08/641.356
; EARLIER FILING DATE: 1996-04-30
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-132-528-3

Query Match 100.0%; Score 1726; DB 3; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
Db 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
Qy 61 PPAPATVPAPPPAAAANTPNACQDPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATVPAPPPAAAANTPNACQDPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPVANDTRIVLGRDOKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPVANDTRIVLGRDOKLYASAEA 180
Qy 181 TDSKAAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSPKNGQIWT 240
Db 181 TDSKAAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSPKNGQIWT 240
Qy 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESTRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESTRPLVAPPAPAPAPAE 300
Qy 301 APAPAGEVAPTPTTPTQRTLPA 325
Db 301 APAPAGEVAPTPTTPTQRTLPA 325

RESULT 5
US-08-875-494-2
; Sequence 2, Application US/08875494
; Patent No. 6221610
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, ANNE
; APPLICANT: MARCHAL, GILLES
; APPLICANT: PESCHER, PASCALE
; APPLICANT: ROMAIN, FELIX
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USES FOR VACCINES AND FOR THE DETECTION OF
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0860-0122-0 PCT
; CURRENT APPLICATION NUMBER: US/08/875.494
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: PCT/FR96/00166

US-09-599-366-2
; Sequence 2, Application US/09599366
; Patent No. 6335181
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/599.366
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 09/132.528
; PRIOR FILING DATE: 1998-08-11
; PRIOR APPLICATION NUMBER: 08/641.356
; PRIOR FILING DATE: 1996-04-30
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-599-366-2

Query Match 100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
Db 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
Qy 61 PPAPATVPAPPPAAAANTPNACQDPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATVPAPPPAAAANTPNACQDPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPVANDTRIVLGRDOKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPVANDTRIVLGRDOKLYASAEA 180
Qy 181 TDSKAAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSPKNGQIWT 240
Db 181 TDSKAAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSPKNGQIWT 240
Qy 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESTRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAESTRPLVAPPAPAPAPAE 300
Qy 301 APAPAGEVAPTPTTPTQRTLPA 325
Db 301 APAPAGEVAPTPTTPTQRTLPA 325
```

Db 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPVPTTAAASPPSTAAA 60
QY 61 PPAPATPVAPPAAANTPNAQPGDPAAPADNAPPPPPVIAAPNAQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPAAANTPNAQPGDPAAPADNAPPPPPVIAAPNAQPVRIIDNPVGGF 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDSKNGQIWT 240
Db 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDSKNGQIWT 240
QY 241 GVIGSPAANADPAGPPQRFVWVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANADPAGPPQRFVWVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAE 300
QY 301 APAPAPAGEVAPTTPTTPTQRTLPA 325
Db 301 APAPAPAGEVAPTTPTTPTQRTLPA 325

RESULT 7
US-09-599-366-3
; Sequence 3, Application US/09599366
; Patent No. 6335181
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/599,366
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: 1998-08-11
; PRIOR APPLICATION NUMBER: 08/641,356
; PRIOR FILING DATE: 1996-04-30
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-599-366-3

Query Match 100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127; Indels 0; Gaps 0;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPVPTTAAASPPSTAAA 60
Db 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPVPTTAAASPPSTAAA 60
QY 61 PPAPATPVAPPAAANTPNAQPGDPAAPADNAPPPPPVIAAPNAQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPAAANTPNAQPGDPAAPADNAPPPPPVIAAPNAQPVRIIDNPVGGF 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDSKNGQIWT 240
Db 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDSKNGQIWT 240
QY 241 GVIGSPAANADPAGPPQRFVWVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANADPAGPPQRFVWVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAE 300

QY 301 APAPAPAGEVAPTTPTTPTQRTLPA 325
Db 301 APAPAPAGEVAPTTPTTPTQRTLPA 325
RESULT 8
US-08-875-494-2
; Sequence 2, Application US/08875494
; Patent No. 6379902
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USES FOR VACCINES AND FOR THE DETECTION OF
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0122-0 PCT
; CURRENT APPLICATION NUMBER: US/08/875,494
; CURRENT FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: PCT/FR96/00166
; PRIOR FILING DATE: 1996-02-01
; PRIOR APPLICATION NUMBER: 382184
; PRIOR FILING DATE: 1995-02-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-08-875-494-2

Query Match 100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127; Indels 0; Gaps 0;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPVPTTAAASPPSTAAA 60
Db 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPVPTTAAASPPSTAAA 60
QY 61 PPAPATPVAPPAAANTPNAQPGDPAAPADNAPPPPPVIAAPNAQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPAAANTPNAQPGDPAAPADNAPPPPPVIAAPNAQPVRIIDNPVGGF 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDSKNGQIWT 240
Db 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDSKNGQIWT 240
QY 241 GVIGSPAANADPAGPPQRFVWVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAE 300
Db 241 GVIGSPAANADPAGPPQRFVWVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAE 300
QY 301 APAPAPAGEVAPTTPTTPTQRTLPA 325
Db 301 APAPAPAGEVAPTTPTTPTQRTLPA 325

RESULT 9
US-09-510-031A-6
; Sequence 6, Application US/09510031A
; Patent No. 6638518
; GENERAL INFORMATION:
; APPLICANT: Ratliff, Timothy
; APPLICANT: Kline, Joel
; TITLE OF INVENTION: METHOD FOR INHIBITING INFLAMMATORY RESPONSES
; FILE REFERENCE: 140.0010 0101
; CURRENT APPLICATION NUMBER: US/09/510,031A
; CURRENT FILING DATE: 2000-02-22

```
; PRIOR APPLICATION NUMBER: US 60/121,177
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-510-031A-6

Query Match          100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDPNLTRKGRGLAALAIAMASASLVTVAVPATANADPEAPPPVPTTAAASPSTAAA 60
DB 1 MHQVDPNLTRKGRGLAALAIAMASASLVTVAVPATANADPEAPPPVPTTAAASPSTAAA 60

QY 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
DB 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120

QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180
DB 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180

QY 181 TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQOIWT 240
DB 181 TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQOIWT 240

QY 241 GVIGSPAANADAGPPQRFVWLTANNPVVDKGAAKALAESIRPLVAPPAPAPAPAE 300
DB 241 GVIGSPAANADAGPPQRFVWLTANNPVVDKGAAKALAESIRPLVAPPAPAPAPAE 300

QY 301 APAPAPAGEVAPTPTTPTPTQRTLPA 325
DB 301 APAPAPAGEVAPTPTTPTPTQRTLPA 325

RESULT 10
US-09-985-372-2
; Sequence 2, Application US/09985372
; Patent No. 6676945
; GENERAL INFORMATION:
; APPLICANT: LAQUEYRERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/985,372
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-3

Query Match          100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDPNLTRKGRGLAALAIAMASASLVTVAVPATANADPEAPPPVPTTAAASPSTAAA 60
DB 1 MHQVDPNLTRKGRGLAALAIAMASASLVTVAVPATANADPEAPPPVPTTAAASPSTAAA 60

QY 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
DB 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120

QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180
DB 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180

QY 181 TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQOIWT 240
DB 181 TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQOIWT 240

QY 241 GVIGSPAANADAGPPQRFVWLTANNPVVDKGAAKALAESIRPLVAPPAPAPAPAE 300
DB 241 GVIGSPAANADAGPPQRFVWLTANNPVVDKGAAKALAESIRPLVAPPAPAPAPAE 300

QY 301 APAPAPAGEVAPTPTTPTPTQRTLPA 325
DB 301 APAPAPAGEVAPTPTTPTPTQRTLPA 325

US-09-985-372-3
; Sequence 3, Application US/09985372
; Patent No. 6676945
; GENERAL INFORMATION:
; APPLICANT: LAQUEYRERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/985,372
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-3

Query Match          100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDPNLTRKGRGLAALAIAMASASLVTVAVPATANADPEAPPPVPTTAAASPSTAAA 60
DB 1 MHQVDPNLTRKGRGLAALAIAMASASLVTVAVPATANADPEAPPPVPTTAAASPSTAAA 60

QY 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
DB 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120

QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180
DB 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180

QY 181 TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQOIWT 240
DB 181 TDSKAAARLGSDMGGEFYMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQOIWT 240

QY 241 GVIGSPAANADAGPPQRFVWLTANNPVVDKGAAKALAESIRPLVAPPAPAPAPAE 300
DB 241 GVIGSPAANADAGPPQRFVWLTANNPVVDKGAAKALAESIRPLVAPPAPAPAPAE 300

QY 301 APAPAPAGEVAPTPTTPTPTQRTLPA 325
DB 301 APAPAPAGEVAPTPTTPTPTQRTLPA 325
```

RESULT 12
US-08-818-112-53
; Sequence 53, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-53

Query Match 100.0%; Score 1726; DB 3; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDNLTTRKGRLLAALAAIAMAASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
DB 8 MHQVDNLTTRKGRLLAALAAIAMAASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 67
QY 61 PPAPATVPVAPPPAAANTNAPQGDPAADPNAPPPVIAAPNAPQPVRIIDNPVGGF 120
DB 68 PPAPATVPVAPPPAAANTNAPQGDPAADPNAPPPVIAAPNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWESDAAHFDYGSALLSKTTGDPDPGPPPPVANDTRI-VLGRDQKLYASAEA 180
DB 128 SFALPAGWESDAAHFDYGSALLSKTTGDPDPGPPPPVANDTRI-VLGRDQKLYASAEA 187
QY 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSKPQNGQIWT 240
DB 188 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSKPQNGQIWT 247
QY 241 GVIGSPAANAPDAGPPQRFVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAP 300
DB 248 GVIGSPAANAPDAGPPQRFVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAP 307
QY 301 APAPAPAGEVAPTPTTPTORTLPA 325

DB 308 APAPAPAGEVAPTPTTPTORTLPA 332

RESULT 13
US-08-818-111-53
; Sequence 53, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-53

Query Match 100.0%; Score 1726; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDNLTTRKGRLLAALAAIAMAASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
DB 8 MHQVDNLTTRKGRLLAALAAIAMAASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 67
QY 61 PPAPATVPVAPPPAAANTNAPQGDPAADPNAPPPVIAAPNAPQPVRIIDNPVGGF 120
DB 68 PPAPATVPVAPPPAAANTNAPQGDPAADPNAPPPVIAAPNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWESDAAHFDYGSALLSKTTGDPDPGPPPPVANDTRI-VLGRDQKLYASAEA 180
DB 128 SFALPAGWESDAAHFDYGSALLSKTTGDPDPGPPPPVANDTRI-VLGRDQKLYASAEA 187
QY 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSKPQNGQIWT 240
DB 188 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDFSKPQNGQIWT 247
QY 241 GVIGSPAANAPDAGPPQRFVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAP 300
DB 248 GVIGSPAANAPDAGPPQRFVWLTGANNPVDKGAALAESIRPLVAPPAPAPAPAP 307
QY 301 APAPAPAGEVAPTPTTPTORTLPA 325

```

Db 308 APAPAPAGEVAPTPTTPTPQRTLPA 332
|||||
RESULT 14
US-09-056-556-53
; Sequence 53, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 622-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-53
Query Match 100.0%; Score 1726; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGRLAALAIAMASASIVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
Db 8 MHQVDPNLTTRKGRLAALAIAMASASIVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
Qy 61 PPAPATPVAPPPPPAAANTPNAQGDPAAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
Db 68 PPAPATPVAPPPPPAAANTPNAQGDPAAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 127
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPQPPVANDTRIVLGRLDQKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPQPPVANDTRIVLGRLDQKLYASAEA 187
Qy 181 TDSKAAARLGSMDGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSPDKPENGQIWT 240
Db 188 TDSKAAARLGSMDGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSPDKPENGQIWT 247
Qy 241 GVIGSPAANADAGPPQRFVWVLGTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 300
Db 248 GVIGSPAANADAGPPQRFVWVLGTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 307
Qy 301 APAPAPAGEVAPTPTTPTPQRTLPA 325
Db 308 APAPAPAGEVAPTPTTPTPQRTLPA 332

```

```

RESULT 15
US-09-072-596-53
; Sequence 53, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-53
Query Match 100.0%; Score 1726; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGRLAALAIAMASASIVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
Db 8 MHQVDPNLTTRKGRLAALAIAMASASIVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
Qy 61 PPAPATPVAPPPPPAAANTPNAQGDPAAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
Db 68 PPAPATPVAPPPPPAAANTPNAQGDPAAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 127
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPQPPVANDTRIVLGRLDQKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPQPPVANDTRIVLGRLDQKLYASAEA 187
Qy 181 TDSKAAARLGSMDGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSPDKPENGQIWT 240
Db 188 TDSKAAARLGSMDGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 247
Qy 241 GVIGSPAANADAGPPQRFVWVLGTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 300
Db 248 GVIGSPAANADAGPPQRFVWVLGTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 307
Qy 301 APAPAPAGEVAPTPTTPTPQRTLPA 325

```


Thu Jul 8 11:11:29 2004

us-10-720-192-2.ra1

Page 8

Db 308 APAPAPAGEVAPTPTTPTQRTLPA 332
|||||

Search completed: July 7, 2004, 18:24:52
Job time : 25.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:24:06 ; Search time 42.5 Seconds
(without alignments)
2380.406 Million cell updates/sec

Title: US-10-720-192-2

Perfect score: 1726

Sequence: 1 MHQVDFNLTRKGRGLAALAI.....PAGEVAPPTPTPTQRTLPA 325

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1726	100.0	325	10	US-09-985-372-2 Sequence 2, Appl
2	1726	100.0	325	10	US-09-985-372-3 Sequence 3, Appl
3	1726	100.0	332	12	US-09-886-349A-41 Sequence 41, Appl
4	1726	100.0	332	14	US-10-193-002-53 Sequence 53, Appl
5	1726	100.0	332	14	US-10-084-843-53 Sequence 53, Appl
6	1726	100.0	332	14	US-10-098-732A-41 Sequence 41, Appl
7	1542	89.3	286	10	US-09-985-372-4 Sequence 4, Appl
8	1499	86.8	652	14	US-10-193-002-350 Sequence 350, App
9	1499	86.8	652	14	US-10-084-843-355 Sequence 355, App
10	1499	86.8	802	9	US-09-287-849-10 Sequence 10, Appl
11	1499	86.8	802	14	US-10-193-002-209 Sequence 209, App
12	1499	86.8	802	14	US-10-193-002-346 Sequence 346, App
13	1499	86.8	802	14	US-10-084-843-214 Sequence 214, App
14	1499	86.8	802	14	US-10-084-843-351 Sequence 351, App
15	1499	86.8	802	14	US-10-359-460-10 Sequence 10, Appl

16	248.5	14.4	273	16	US-10-437-963-184760	Sequence 184760,
17	241	13.0	402	16	US-10-437-963-141313	Sequence 141313,
18	240.5	13.9	437	16	US-10-437-963-168762	Sequence 168762,
19	239	13.8	431	16	US-10-437-963-204963	Sequence 204963,
20	238	13.8	1369	16	US-10-437-963-190985	Sequence 190985,
21	237.5	13.8	547	12	US-10-425-114-41545	Sequence 41545, A
22	237.5	13.5	597	16	US-10-437-963-102579	Sequence 102579,
23	232.5	13.5	270	16	US-10-437-963-130068	Sequence 130068,
24	232.5	13.5	360	16	US-10-437-963-179473	Sequence 179473,
25	232.5	13.4	426	16	US-10-437-963-142229	Sequence 142229,
26	230.5	13.4	584	14	US-10-156-761-12405	Sequence 12405, A
27	229.5	13.3	569	16	US-10-437-963-149276	Sequence 149276,
28	229	13.3	440	16	US-10-437-963-128104	Sequence 128104,
29	228.5	13.2	235	16	US-10-437-963-162861	Sequence 162861,
30	227.5	13.2	652	15	US-10-104-047-3364	Sequence 3364, Ap
31	227	13.2	598	12	US-10-282-122A-64954	Sequence 64954, A
32	226.5	13.1	504	16	US-10-437-963-128105	Sequence 128105,
33	223.5	12.9	299	16	US-10-437-963-134054	Sequence 134054,
34	223.5	12.9	501	16	US-10-437-963-161137	Sequence 161137,
35	223.5	12.9	533	16	US-10-437-963-167716	Sequence 167716,
36	221.5	12.8	556	16	US-10-437-963-137918	Sequence 137918,
37	220.5	12.8	406	16	US-10-437-963-122256	Sequence 122256,
38	220	12.7	389	12	US-10-425-114-56400	Sequence 56400, A
39	220	12.7	466	16	US-10-437-963-195119	Sequence 195119,
40	219.5	12.7	369	16	US-10-437-963-196554	Sequence 196554,
41	219.5	12.7	376	16	US-10-437-963-154970	Sequence 154970,
42	218.5	12.7	285	16	US-10-437-963-189482	Sequence 189482,
43	218.5	12.7	534	16	US-10-437-963-159895	Sequence 159895,
44	217.5	12.6	350	16	US-10-437-963-102583	Sequence 102583,
45	217.5	12.6	817	15	US-10-369-493-1813	Sequence 1813, Ap

ALIGNMENTS

RESULT 1

US-09-985-372-2
; Sequence 2, Application US/09985372
; Publication No. US20030054008A1
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/985,372
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-2

Query Match 100.0%; Score 1726; DB 10; Length 325;

Best Local Similarity 100.0%; Pred. No. 8.6e+106; Mismatches 0; Indels 0; Gaps 0;

Matches 325; Conservative 0;

Qy	1	MHQVDFNLTRKGRGLAALAI	AAMASASLVTVAVPATANADPEPAPVPTTAA	SPSTAAA	60
Db	1	MHQVDFNLTRKGRGLAALAI	AAMASASLVTVAVPATANADPEPAPVPTTAA	SPSTAAA	60
Qy	61	PPAPATPVAPPPAAANTENAQ	GGPNAAPPADNAPPPVIA	NPQVRI	120
Db	61	PPAPATPVAPPPAAANTENAQ	GGPNAAPPADNAPPPVIA	NPQVRI	120
Qy	121	SPALPAGWVESDAAHFDYGSALLSKT	TGDPFPFGPPVANDTRIVLGR	LQKLYASAE	180

Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIVLGRDQKLYASAEA 180
Qy 181 TDSKAAARLGSDMGDFYFYPGTRINQETVSLDANGVSGSASYEYVFXFSDPSKNGQIWT 240
Db 181 TDSKAAARLGSDMGDFYFYPGTRINQETVSLDANGVSGSASYEYVFXFSDPSKNGQIWT 240
Qy 241 GVIGSPAANAPDAGPPQRFVWLTANNPVDKGAALAESIRPLVAPAPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPPQRFVWLTANNPVDKGAALAESIRPLVAPAPAPAPAPAE 300
Qy 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
Db 301 APAPAPAGEVAPTPTPTPTQRTPLA 325

RESULT 2
US-09-985-372-3
; Sequence 3, Application US/09985372
; Publication No. US20030054008A1
; GENERAL INFORMATION:
; APPLICANT: LAQUEYRERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/985,372
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-3

Query Match 100.0%; Score 1726; DB 10; Length 325;
Best Local Similarity 100.0%; Pred. No. 8.6e-106;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPPVPTTAASPPSTAAA 60
Db 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPPVPTTAASPPSTAAA 60
Qy 61 PPAPATPVAPPPAAAANTPNAQPGDPNAAAPPADPNAPPPVIAFNAPQVRIIDNPVGGF 120
Db 61 PPAPATPVAPPPAAAANTPNAQPGDPNAAAPPADPNAPPPVIAFNAPQVRIIDNPVGGF 120
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIVLGRDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIVLGRDQKLYASAEA 180
Qy 181 TDSKAAARLGSDMGDFYFYPGTRINQETVSLDANGVSGSASYEYVFXFSDPSKNGQIWT 240
Db 181 TDSKAAARLGSDMGDFYFYPGTRINQETVSLDANGVSGSASYEYVFXFSDPSKNGQIWT 240
Qy 241 GVIGSPAANAPDAGPPQRFVWLTANNPVDKGAALAESIRPLVAPAPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPPQRFVWLTANNPVDKGAALAESIRPLVAPAPAPAPAPAE 300
Qy 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
Db 301 APAPAPAGEVAPTPTPTPTQRTPLA 325

RESULT 3
US-09-886-349A-41
; Sequence 41, Application US/09986349A

; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: DPEP
US-09-886-349A-41

Query Match 100.0%; Score 1726; DB 12; Length 332;
Best Local Similarity 100.0%; Pred. No. 8.7e-106;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPPVPTTAASPPSTAAA 60
Db 8 MHQVDNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPPPVPTTAASPPSTAAA 67
Qy 61 PPAPATPVAPPPAAAANTPNAQPGDPNAAAPPADPNAPPPVIAFNAPQVRIIDNPVGGF 120
Db 68 PPAPATPVAPPPAAAANTPNAQPGDPNAAAPPADPNAPPPVIAFNAPQVRIIDNPVGGF 127
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIVLGRDQKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIVLGRDQKLYASAEA 187
Qy 181 TDSKAAARLGSDMGDFYFYPGTRINQETVSLDANGVSGSASYEYVFXFSDPSKNGQIWT 240
Db 188 TDSKAAARLGSDMGDFYFYPGTRINQETVSLDANGVSGSASYEYVFXFSDPSKNGQIWT 247
Qy 241 GVIGSPAANAPDAGPPQRFVWLTANNPVDKGAALAESIRPLVAPAPAPAPAPAE 300
Db 248 GVIGSPAANAPDAGPPQRFVWLTANNPVDKGAALAESIRPLVAPAPAPAPAPAE 307
Qy 301 APAPAPAGEVAPTPTPTPTQRTPLA 325
Db 308 APAPAPAGEVAPTPTPTPTQRTPLA 332

RESULT 4
US-10-193-002-53
; Sequence 53, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedgick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/193,002
 FILING DATE: 10-Jul-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,596
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C9
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 53:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 332 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 53:
 US-10-193-002-53

Query Match 100.0%; Score 1726; DB 14; Length 332;
 Best Local Similarity 100.0%; Pred. No. 8.7e-106;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEQVDNLTREKRLAALAIAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 60
 DB 8 MEQVDNLTREKRLAALAIAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 67
 QY 61 PPAPATPVAPPVPPAAANTPNAQPDGDNAAAPPADPNAPPPVIAAPNAPQVRIDNPVGGF 120
 DB 68 PPAPATPVAPPVPPAAANTPNAQPDGDNAAAPPADPNAPPPVIAAPNAPQVRIDNPVGGF 127
 QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPPVANDTRIVLGRDQKLYASAEA 180
 DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPPVANDTRIVLGRDQKLYASAEA 187
 QY 181 TDSKAAARLGSDMGGEFYMPYGTINQETVSLDANGVSGSASYEYVKFSPKNGQIWT 240
 DB 188 TDSKAAARLGSDMGGEFYMPYGTINQETVSLDANGVSGSASYEYVKFSPKNGQIWT 247
 QY 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVVDKGAALAESIRPLVAPPAPAPAPAP 300
 DB 248 GVIGSPAANAPDAGPPQRFVWVLTANNPVVDKGAALAESIRPLVAPPAPAPAPAP 307
 QY 301 APAPAPAGEVAPTPTTPTPQRTLPA 325
 DB 308 APAPAPAGEVAPTPTTPTPQRTLPA 332

RESULT 5
 US-10-084-843-53
 ; Sequence 53, Application US/10084843
 ; Publication No. US20030143243A1
 ; GENERAL INFORMATION:
 APPLICANT: Reed, Steven G.
 Skeiky, Yasir A.W.
 Dillon, David C.
 Campos-Neto, Antonio
 Houghton, Raymond
 Vedwick, Thomas S.
 Twardzik, Daniel R.

Lodes, Michael J.
 Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 AND DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 355
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/084,843
 FILING DATE: 25-Feb-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 53:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 332 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 53:
 US-10-084-843-53

Query Match 100.0%; Score 1726; DB 14; Length 332;
 Best Local Similarity 100.0%; Pred. No. 8.7e-106;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEQVDNLTREKRLAALAIAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 60
 DB 8 MEQVDNLTREKRLAALAIAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 67
 QY 61 PPAPATPVAPPVPPAAANTPNAQPDGDNAAAPPADPNAPPPVIAAPNAPQVRIDNPVGGF 120
 DB 68 PPAPATPVAPPVPPAAANTPNAQPDGDNAAAPPADPNAPPPVIAAPNAPQVRIDNPVGGF 127
 QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPPVANDTRIVLGRDQKLYASAEA 180
 DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPPVANDTRIVLGRDQKLYASAEA 187
 QY 181 TDSKAAARLGSDMGGEFYMPYGTINQETVSLDANGVSGSASYEYVKFSPKNGQIWT 240
 DB 188 TDSKAAARLGSDMGGEFYMPYGTINQETVSLDANGVSGSASYEYVKFSPKNGQIWT 247
 QY 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVVDKGAALAESIRPLVAPPAPAPAPAP 300
 DB 248 GVIGSPAANAPDAGPPQRFVWVLTANNPVVDKGAALAESIRPLVAPPAPAPAPAP 307
 QY 301 APAPAPAGEVAPTPTTPTPQRTLPA 325
 DB 308 APAPAPAGEVAPTPTTPTPQRTLPA 332

RESULT 6
 US-10-098-732A-41
 ; Sequence 41, Application US/10098732A
 ; Publication No. US20030175294A1

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DPEP
US-10-098-732A-41

Query Match 100.0%; Score 1726; DB 14; Length 332;
Best Local Similarity 100.0%; Pred. No. 8.7e-106;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDPNLTERRKRLAALAIAMASASLVTVAVPATANADPEAPPVPTTAASPPSTAAA 60
DB 8 MHQVDPNLTERRKRLAALAIAMASASLVTVAVPATANADPEAPPVPTTAASPPSTAAA 67

QY 61 PPAPATPVAPPPAAANTPNAQGDPAAPPADPNAPPPVPIAPNAPQPVRIINPVGGF 120
DB 68 PPAPATPVAPPPAAANTPNAQGDPAAPPADPNAPPPVPIAPNAPQPVRIINPVGGF 127

QY 121 SPALPAGWVESDAARHFDYGSALLSKTTGDPFPFGQPPVANDTRIVLGRDQKLYASAEA 180
DB 128 SPALPAGWVESDAARHFDYGSALLSKTTGDPFPFGQPPVANDTRIVLGRDQKLYASAEA 187

QY 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDPSPKNGQIWT 240
DB 188 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEVKFSDPSPKNGQIWT 247

QY 241 GVIGSPAANAPDAGPPQRFVFWLGTANNPVDKGAALASIRPLVAPPPAPAPAPAP 300
DB 248 GVIGSPAANAPDAGPPQRFVFWLGTANNPVDKGAALASIRPLVAPPPAPAPAPAP 307

QY 301 APAPAPAGEVAPTPTPTPQRTLPA 325
DB 308 APAPAPAGEVAPTPTPTPQRTLPA 332

RESULT 7
US-09-985-372-4
; Sequence 4, Application US/09985372
; Publication No. US20030054008A1
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICRORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/985,372
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 286
; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-4

Query Match 89.3%; Score 1542; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 9.3e-94;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 DPEPAPPVPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQGDPAAPPADPNAPP 99
DB 1 DPEPAPPVPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQGDPAAPPADPNAPP 60

QY 100 PVIAENAPQPVRIINPVGGFSAFPAGWVESDAARHFDYGSALLSKTTGDPFPFGQPPV 159
DB 61 PVIAENAPQPVRIINPVGGFSAFPAGWVESDAARHFDYGSALLSKTTGDPFPFGQPPV 120

QY 160 ANDTRIVLGRDQKLYASAEATDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSG 219
DB 121 ANDTRIVLGRDQKLYASAEATDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSG 180

QY 220 SASYYEVKFSDPSPKNGQIWTGVIGSPAANAPDAGPPQRFVFWLGTANNPVDKGAAL 279
DB 181 SASYYEVKFSDPSPKNGQIWTGVIGSPAANAPDAGPPQRFVFWLGTANNPVDKGAAL 240

QY 280 AESIRPLVAPPPAP 325
DB 241 AESIRPLVAPPPAP 286

RESULT 8
US-10-193-002-350
; Sequence 350, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, David C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 350:
; SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 652 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 350:
US-10-193-002-350

Query Match      86.8%; Score 1499; DB 14; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.5e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPAAANTNAQPGDPAAPADPNAPPVTAAPN 106
DB 374 VPTTAASPPSTAAAPAPATPVAPPAAANTNAQPGDPAAPADPNAPPVTAAPN 433
QY 107 APQVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 166
DB 434 APQVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 493
QY 167 LGRLDQKLYASAEATDSKAAARLGSMDGFEFFYMPYPGTRINQETVSLDANGVSGSASYEV 226
DB 494 LGRLDQKLYASAEATDSKAAARLGSMDGFEFFYMPYPGTRINQETVSLDANGVSGSASYEV 553
QY 227 KFSDSPKNGQIWTGIVGSPAANADAGPPQRFVWVLGTANNPVVDKGAALKALAESIRPL 286
DB 554 KFSDSPKNGQIWTGIVGSPAANADAGPPQRFVWVLGTANNPVVDKGAALKALAESIRPL 613
QY 287 VAPPAPAPAPAEAPAPAPAGEVAPTPTTPTPQRTLPA 325
DB 614 VAPPAPAPAPAEAPAPAPAGEVAPTPTTPTPQRTLPA 652

RESULT 9
US-10-084-843-355
; Sequence 355, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Applicant: Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
```

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 355:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 355:
US-10-084-843-355

Query Match      86.8%; Score 1499; DB 14; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.5e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPAAANTNAQPGDPAAPADPNAPPVTAAPN 106
DB 374 VPTTAASPPSTAAAPAPATPVAPPAAANTNAQPGDPAAPADPNAPPVTAAPN 433
QY 107 APQVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 166
DB 434 APQVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 493
QY 167 LGRLDQKLYASAEATDSKAAARLGSMDGFEFFYMPYPGTRINQETVSLDANGVSGSASYEV 226
DB 494 LGRLDQKLYASAEATDSKAAARLGSMDGFEFFYMPYPGTRINQETVSLDANGVSGSASYEV 553
QY 227 KFSDSPKNGQIWTGIVGSPAANADAGPPQRFVWVLGTANNPVVDKGAALKALAESIRPL 286
DB 554 KFSDSPKNGQIWTGIVGSPAANADAGPPQRFVWVLGTANNPVVDKGAALKALAESIRPL 613
QY 287 VAPPAPAPAPAEAPAPAPAGEVAPTPTTPTPQRTLPA 325
DB 614 VAPPAPAPAPAEAPAPAPAGEVAPTPTTPTPQRTLPA 652

RESULT 10
US-09-287-849-10
; Sequence 10, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Applicant: Skeiky, Yasir A.W.
; Dillon, Davin C.
; Applicant: Alderson, Mark
; Applicant: Campos-Neto, Antonio
; Applicant: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/816,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10
```

Query Match 86.8%; Score 1499; DB 9; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPNAPPPADPNAPPPVIAPN 106
DB 524 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPNAPPPADPNAPPPVIAPN 583

QY 107 APOQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPVANDTRIV 166
DB 584 APOQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFGQPPPVANDTRIV 643

QY 167 LGRLDQKLYASAEATDSKAAARLGSDMGFEFYMYPGTRINQETVSLDANGVSGSASYEV 226
DB 644 LGRLDQKLYASAEATDSKAAARLGSDMGFEFYMYPGTRINQETVSLDANGVSGSASYEV 703

QY 227 KFSDDSKPQNGQIWTGIVGSPAANADPAGPQRFVWVWLTGANNPVDKGAALAESIRPL 286
DB 704 KFSDDSKPQNGQIWTGIVGSPAANADPAGPQRFVWVWLTGANNPVDKGAALAESIRPL 763

QY 287 VAPPPAPAPAPAPAPAPAPAGEVAPTPTTTPQRTPLA 325
DB 764 VAPPPAPAPAPAPAPAPAPAGEVAPTPTTTPQRTPLA 802

RESULT 12
US-10-193-002-346
; Sequence 346, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedrick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 209:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid


```
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 802 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 346:
US-10-193-002-346

Query Match      86.8%; Score 1499; DB 14; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPAAANTPNAQPGDPNAAPPADPNAPPVPIAPN 106
DB 524 VPTTAASPPSTAAAPAPATPVAPPAAANTPNAQPGDPNAAPPADPNAPPVPIAPN 583
QY 107 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPVANDTRIV 166
DB 584 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPVANDTRIV 643
QY 167 LGRLDQKLYASAEATDSKAAARLGSDMGEFYMPYPTGTRINQETVSLDANGVSGSASYEV 226
DB 644 LGRLDQKLYASAEATDSKAAARLGSDMGEFYMPYPTGTRINQETVSLDANGVSGSASYEV 703
QY 227 KFSDPKSKNGQIWTGVI GSPAAAPDAGPPQRFVFWVLGTANNPVVDKGAALAEIRPL 286
DB 704 KFSDPKSKNGQIWTGVI GSPAAAPDAGPPQRFVFWVLGTANNPVVDKGAALAEIRPL 286
QY 287 VAPPAPAPAPAPAPAPAPAGEVAPTTPTTPTQRTPLA 325
DB 764 VAPPAPAPAPAPAPAPAPAGEVAPTTPTTPTQRTPLA 802

RESULT 13
US-10-084-843-214
; Sequence 214, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
;   Skeiky, Yasir A.W.
;   Dillon, Davin C.
;   Campos-Neto, Antonio
;   Houghton, Raymond
;   Vedvick, Thomas S.
;   Twardzik, Daniel R.
;   Lodes, Michael J.
;   Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 802 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 346:
US-10-193-002-346

Query Match      86.8%; Score 1499; DB 14; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPAAANTPNAQPGDPNAAPPADPNAPPVPIAPN 106
DB 524 VPTTAASPPSTAAAPAPATPVAPPAAANTPNAQPGDPNAAPPADPNAPPVPIAPN 583
QY 107 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPVANDTRIV 166
DB 584 APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPPVANDTRIV 643
QY 167 LGRLDQKLYASAEATDSKAAARLGSDMGEFYMPYPTGTRINQETVSLDANGVSGSASYEV 226
DB 644 LGRLDQKLYASAEATDSKAAARLGSDMGEFYMPYPTGTRINQETVSLDANGVSGSASYEV 703
QY 227 KFSDPKSKNGQIWTGVI GSPAAAPDAGPPQRFVFWVLGTANNPVVDKGAALAEIRPL 286
DB 704 KFSDPKSKNGQIWTGVI GSPAAAPDAGPPQRFVFWVLGTANNPVVDKGAALAEIRPL 286
QY 287 VAPPAPAPAPAPAPAPAPAGEVAPTTPTTPTQRTPLA 325
DB 764 VAPPAPAPAPAPAPAPAPAGEVAPTTPTTPTQRTPLA 802

RESULT 14
US-10-084-843-351
; Sequence 351, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
;   Skeiky, Yasir A.W.
;   Dillon, Davin C.
;   Campos-Neto, Antonio
;   Houghton, Raymond
;   Vedvick, Thomas S.
;   Twardzik, Daniel R.
;   Lodes, Michael J.
;   Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
```

APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/POCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match 86.8%; Score 1499; DB 14; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	47	VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPNAAAPPADPNAPPVIAPN	106
DB	524	VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPNAAAPPADPNAPPVIAPN	583
QY	107	APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPVANDTRIV	166
DB	584	APQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPVANDTRIV	643
QY	167	LGRLDQKLYASAEATDSKAAARLGSDMGCFYMPYPGTRINQETVSLDANGVSGSASYEV	226
DB	644	LGRLDQKLYASAEATDSKAAARLGSDMGCFYMPYPGTRINQETVSLDANGVSGSASYEV	703
QY	227	KFSDPSKNGQIWTGVIQSPAAADPAGPPQRFVWVLGTANNPVDKGAALKALAESIRPL	286
DB	704	KFSDPSKNGQIWTGVIQSPAAADPAGPPQRFVWVLGTANNPVDKGAALKALAESIRPL	763
QY	287	VAPPAPAPAPAPAPAPAPAGEVAPTPTTPTPQRTLPA	325
DB	764	VAPPAPAPAPAPAPAPAPAGEVAPTPTTPTPQRTLPA	802

Search completed: July 7, 2004, 18:33:33
Job time : 44.5 secs

RESULT 15
US-10-359-460-10
Sequence 10, Application US/10359460
Publication No. US20030147911A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:17:14 ; Search time 50.5 seconds
(without alignments)
1818.375 Million cell updates/sec

Title: us-10-720-192-3

Perfect score: 1726

Sequence: 1 MHQVDPNLTFRKGRILAALAI.....PAGEVAPTTPTPTORTLPA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1726	100.0	325	AAW03565	AAW03565 Mycobacte
2	1726	100.0	325	AAW0732	AAW0732 Mycobacte
3	1726	100.0	332	AAW32350	AAW32350 Mycobacte
4	1726	100.0	332	AAW32418	AAW32418 Mycobacte
5	1726	100.0	332	AAW64322	AAW64322 Mycobacte
6	1726	100.0	332	AAW81683	AAW81683 M. tuberc
7	1726	100.0	332	AAW38945	AAW38945 M. tuberc
8	1726	100.0	332	AAW39083	AAW39083 M. tuberc
9	1726	100.0	332	AAW01895	AAW01895 M. tuberc
10	1726	100.0	332	AAE29720	AAE29720 Mycobacte
11	1726	100.0	332	AAE17584	AAE17584 Mycobacte
12	1720	99.7	325	AAW97281	AAW97281 Fibronect
13	1542	89.3	286	AAW03566	AAW03566 Mycobacte
14	1499	86.8	652	AAW39082	AAW39082 M. tubercu
15	1499	86.8	652	AAW39225	AAW39225 M. tuberc
16	1499	86.8	802	AAW64379	AAW64379 Mycobacte
17	1499	86.8	802	AAW81746	AAW81746 M. tuberc
18	1499	86.8	802	AAW32063	AAW32063 Mycobacte
19	1499	86.8	802	AAW39033	AAW39033 M. tuberc
20	1499	86.8	802	AAW39081	AAW39081 M. tubercu
21	1499	86.8	802	AAW39176	AAW39176 M. tuberc
22	1499	86.8	802	AAW39224	AAW39224 M. tuberc
23	1499	86.8	802	AAW74592	AAW74592 Antigenic
24	980	56.8	287	ADB74347	ADB74347 Mycobacte
25	237.5	13.8	268	ABU52942	ABU52942 Human mam

26	235	13.6	256	4	ABU52945	Human mam
27	233.5	13.5	262	4	ABU52944	Human mam
28	233	13.5	270	4	ABU52939	Human mam
29	233	13.5	8991	6	ABU08487	S. pneumo
30	228	13.2	544	3	AAG15453	Arabidops
31	227.5	13.2	652	7	ADB65210	Human pro
32	227	13.2	598	6	ABU37030	Protein e
33	225.5	13.1	446	4	ABB70063	Herbicida
34	223.5	12.9	652	5	ABB92424	Herbicida
35	222.5	12.9	267	4	ABU52938	Human mam
36	222	12.9	763	2	AAW31852	Mycobacte
37	221.5	12.8	260	4	ABU52932	Human mam
38	221.5	12.8	267	4	ABU52931	Human mam
39	218.5	12.7	253	4	ABU52937	Human mam
40	217.5	12.6	817	6	ABR53281	Protein s
41	217	12.6	261	4	ABU52940	Human mam
42	217	12.6	277	4	ABU52936	Human mam
43	216.5	12.5	254	4	ABU52933	Human mam
44	214	12.4	406	4	ABG27250	Novel hum
45	212.5	12.3	258	4	ABU52941	Human mam

ALIGNMENTS

RESULT 1
AAW03565
ID AAW03565 standard; protein; 325 AA.

XX AAW03565;

DT 22-APR-1997 (first entry)

DE Mycobacterium tuberculosis antigenic determinant protein.

XX Antigenic determinant; Mycobacterium tuberculosis; ion exchange; human; chromatography; gel filtration; reverse phase column chromatography; immunogenic; serum; guinea pig; expression vector; cosmid; antibody; Mycobacterium smegmatis; Mycobacterium bovis BCG; microorganism; vaccine; hybrid; epitope; disease; diphtheria; cholera; toxin.

OS Mycobacterium tuberculosis.

FH Key Location/Qualifiers
FT Peptide 1..39
FT /note= "signal peptide"
FT Protein 40..325
FT /note= "mature protein"

XX WO9623985-A1.

XX 08-AUG-1996.

XX 31-JAN-1996; 96WO-FR000166.

XX 01-FEB-1995; 95US-00362184.

XX (INSP) INST PASTEUR.

XX Laqueyrie A, Marchal G, Pescher P, Romain F;

XX WPI; 1996-371433/37.

XX N-PSDB; AAT39357.

XX Mycobacterium tuberculosis antigen and hybrid proteins comprising it - useful in vaccines against tuberculosis and in immunoassays.

XX Claim 2; Page 49-50; 74pp; French.

CC This is the amino acid sequence of an antigenic determinant protein from Mycobacterium tuberculosis. The mature protein has calculated mol. wt. of 28779 Da but has an experimental mol. wt of 45-47 kD as determined by SDS -PAGE. The difference is thought to be due to the high frequency of Pro


```

XW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
XX
PR 22-SEP-1995; 95US-00532136.
XX
PR 22-MAR-1996; 96US-00620280.
XX
PR 05-JUN-1996; 96US-00658800.
XX
PR 12-JUL-1996; 96US-00680573.
XX
PA (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX
XX WPI; 1997-192904/17.
DR N-PSDB; AAT91423.
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX useful for diagnosis of M. tuberculosis infection.
XX
XX Example 1; Page 92-94; 190pp; English.
XX
XX A new immunogenic polypeptide has been developed comprising an
XX immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
XX variant differing only in conservative substitutions and/or
XX modifications). The present sequence represents a M.tuberculosis antigen,
XX DPEP. The immunogenic polypeptide can be used to diagnose M.tuberculosis
XX infection by forming complexes with specific antibodies in the sample.
XX Fragments of DNA encoding the immunogenic polypeptide can be used as
XX diagnostic primers or probes and agents that bind to the antigen.
XX especially monoclonal antibodies or equivalent polyclonal antibodies, are
XX also used for diagnosis
XX
XX Sequence 332 AA;
XX
XX Query Match 100.0%; Score 1726; DB 2; Length 332;
XX Best Local Similarity 100.0%; Pred. No. 1.3e-104;
XX Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MHQVDPNLTRKGRLLAALAIAMASASLVTVAVPATANADPEAPVPTTAAASPTAAA 60
DB 8 MHQVDPNLTRKGRLLAALAIAMASASLVTVAVPATANADPEAPVPTTAAASPTAAA 67
XX
QY 61 PPAPATPVAPPPPPAAANTPNAQGDENAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
DB 68 PPAPATPVAPPPPPAAANTPNAQGDENAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 127
XX
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180
DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRDLQKLYASAEA 187
XX
QY 181 TDSKAAARLGSDMGGEFYMPYPGTINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240
DB 188 TDSKAAARLGSDMGGEFYMPYPGTINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 247
XX
QY 241 GVIGSPAANADAGPPQRFVFWVLGTANNPVDGAAKALAESIRPLVAPPAPAPAPAE 300
DB 248 GVIGSPAANADAGPPQRFVFWVLGTANNPVDGAAKALAESIRPLVAPPAPAPAPAE 307
XX
QY 301 APAPAPAGEVAPTETTTPTORTLPA 325
DB 308 APAPAPAGEVAPTETTTPTORTLPA 332
XX
RESULT 4
AAW32418

```

```

AAW32418 standard; protein; 332 AA.
AAW32418;
08-JAN-1998 (first entry)
Mycobacterium tuberculosis antigen DPEP.
Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
skin testing; M.tuberculosis.
Mycobacterium tuberculosis.
WO9709428-A2.
13-MAR-1997.
30-AUG-1996; 96WO-US014674.
01-SEP-1995; 95US-00523436.
22-SEP-1995; 95US-00533634.
22-MAR-1996; 96US-00620874.
05-JUN-1996; 96US-00659683.
12-JUL-1996; 96US-00680574.
(CORI-) CORIXA CORP.
Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
Vedvick TH, Twardzik DR;
WPI; 1997-192903/17.
N-PSDB; AAT91486.
New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
useful in vaccines for prevention or treatment of tuberculosis, also for
diagnosis.
Example 1; Page 87-89; 168pp; English.
A new immunogenic polypeptide has been developed comprising an
immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
variant differing only in conservative substitutions and/or
modifications). The present sequence represents a M.tuberculosis antigen,
DPEP. The immunogenic protein, and fusion proteins containing one or more
of the proteins or one of the proteins plus ESAT-6, are useful in
vaccines, preferably when formulated with a non-specific adjuvant, to
induce an immune response against M.tuberculosis (for treatment or
prevention)
Sequence 332 AA;
Query Match 100.0%; Score 1726; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.3e-104;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDPNLTRKGRLLAALAIAMASASLVTVAVPATANADPEAPVPTTAAASPTAAA 60
DB 8 MHQVDPNLTRKGRLLAALAIAMASASLVTVAVPATANADPEAPVPTTAAASPTAAA 67
XX
QY 61 PPAPATPVAPPPPPAAANTPNAQGDENAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
DB 68 PPAPATPVAPPPPPAAANTPNAQGDENAPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 127
XX
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRDLQKLYASAEA 180
DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRDLQKLYASAEA 187
XX
QY 181 TDSKAAARLGSDMGGEFYMPYPGTINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240
DB 188 TDSKAAARLGSDMGGEFYMPYPGTINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 247
XX
QY 241 GVIGSPAANADAGPPQRFVFWVLGTANNPVDGAAKALAESIRPLVAPPAPAPAPAE 300

```

Db 248 GVIGSPAANADAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPPAPAPAPAP 307
 QY 301 APAPAPAGEVAPTPTTPTPTPTPT 325
 Db 308 APAPAPAGEVAPTPTTPTPTPTPT 332

RESULT 5
 AAW64322
 ID AAW64322 standard; protein; 332 AA.
 XX
 AC AAW64322;
 XX
 DT 17-OCT-2003 (revised)
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen DPEP.
 XX
 DE Tuberculosis; infection; diagnosis; antigen; DPEP.
 XX
 KW Mycobacterium tuberculosis; strain H37Rv.
 XX
 OS
 XX
 FN W09816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018214.
 XX
 PR 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00818111.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 DR WPI; 1998-251292/22.
 DR N-PSDB; AAV44390.
 XX
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 PS Example 1; Page 93-94; 250pp; English.
 XX
 CC This polypeptide comprises the Mycobacterium tuberculosis antigen DPEP. A
 CC DNA sequence (see AAV44390) coding for DPEP was isolated from a M.
 CC tuberculosis genomic library using a probe based on an isolated N-
 CC terminal peptide (see AAW64379). The invention relates to compositions
 CC and methods for diagnosing tuberculosis. It provides polypeptides (see
 CC AAW64291-W64379) comprising an antigenic portion of a soluble M.
 CC tuberculosis antigen, or an immunogenic portion of an M. tuberculosis
 CC antigen, as well as DNA sequences encoding such polypeptides, recombinant
 CC expression vectors and transformed or transfected host cells. Also
 CC claimed are methods and diagnostic kits for detecting M. tuberculosis
 CC infection in a patient using these polypeptides, antibodies or
 CC oligonucleotide probes and primers, for the diagnosis of tuberculosis.
 CC (Updated on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 2; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e-104;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPVPTTAAASPTAAA 60
 Db 8 MHQVDNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPVPTTAAASPTAAA 67
 QY 61 PPAPATVPAPPPAAAANTNAPGDPNAAAPPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
 Db 68 PPAPATVPAPPPAAAANTNAPGDPNAAAPPADPNAPPPVIAFNAPQPVRIIDNPVGGF 127

QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVANDTRIVLGRLOKLYASAEA 180
 Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPPGQPPVANDTRIVLGRLOKLYASAEA 187
 QY 181 TDSKAARLIGSDMGREFYMPGTRINQETVSLDANGVSSASYEVKFSKPKNGQIWT 240
 Db 188 TDSKAARLIGSDMGREFYMPGTRINQETVSLDANGVSSASYEVKFSKPKNGQIWT 247
 QY 241 GVIGSPAANADAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPPAPAPAPAP 300
 Db 248 GVIGSPAANADAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPPAPAPAPAP 307
 QY 301 APAPAPAGEVAPTPTTPTPTPTPT 325
 Db 308 APAPAPAGEVAPTPTTPTPTPTPT 332

RESULT 6
 AAW81683
 ID AAW81683 standard; protein; 332 AA.
 XX
 AC AAW81683;
 XX
 DT 27-JAN-1999 (first entry)
 DT
 XX
 DE M. tuberculosis immunogenic polypeptide DPEP.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FN W09816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018293.
 XX
 PR 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 DR WPI; 1998-261042/23.
 DR N-PSDB; AAV64498.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 PS Example 1; Page 91-92; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX
 SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 2; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e-104;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPVPTTAAASPTAAA 60
 Db 8 MHQVDNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPVPTTAAASPTAAA 67
 QY 61 PPAPATVPAPPPAAAANTNAPGDPNAAAPPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120

Db 68 PPAPATPVAPPPPPAAAANTPNAQPGDPAAPPPADPNAPPPVIAFNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDOKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDOKLYASAEA 187
QY 181 TDSKAAARLGSDMGFEFMPYPTGTRINQETVSLDANGVSGSASYEYKFSKPNQIWT 240
Db 188 TDSKAAARLGSDMGFEFMPYPTGTRINQETVSLDANGVSGSASYEYKFSKPNQIWT 247
QY 241 GVIGSPAANAPDAGPPQRFVWVWLTANNPVDKGAALAEIRPLVAPPAPAPAPAE 300
Db 248 GVIGSPAANAPDAGPPQRFVWVWLTANNPVDKGAALAEIRPLVAPPAPAPAE 307
QY 301 APAPAPAGEVAPTPTTPTPQRTPLA 325
Db 308 APAPAPAGEVAPTPTTPTPQRTPLA 332

RESULT 7
AAY38945
ID AAY38945 standard; protein; 332 AA.
XX AC AAY38945;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis recombinant antigen protein DPEP.
XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
XX KW vaccine; immunity.
XX OS Mycobacterium tuberculosis.
XX PN WO9942118-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US003265.
XX PR 18-FEB-1998; 98US-00024753.
XX PR 05-MAY-1998; 98US-00072596.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX DR WPI; 1999-527416/44.
XX DR N-PSDB; AAZ19088.
XX PT New polypeptide comprising antigenic portions of M. tuberculosis.
XX PS Example 1; Page 131-132; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against M.
CC tuberculosis infection. The new detection methods are needed as current
CC vaccination strategies do not provide 100% immunity
XX SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. NO. 1.3e-104;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDPNLTNRKGRGLAALAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
Db 8 MHQVDPNLTNRKGRGLAALAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 67

QY 61 PPAPATPVAPPPPPAAAANTPNAQPGDPAAPPPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
Db 68 PPAPATPVAPPPPPAAAANTPNAQPGDPAAPPPADPNAPPPVIAFNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDOKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGQPPPVANDTRIVLGRDOKLYASAEA 187
QY 181 TDSKAAARLGSDMGFEFMPYPTGTRINQETVSLDANGVSGSASYEYKFSKPNQIWT 240
Db 188 TDSKAAARLGSDMGFEFMPYPTGTRINQETVSLDANGVSGSASYEYKFSKPNQIWT 247
QY 241 GVIGSPAANAPDAGPPQRFVWVWLTANNPVDKGAALAEIRPLVAPPAPAPAE 300
Db 248 GVIGSPAANAPDAGPPQRFVWVWLTANNPVDKGAALAEIRPLVAPPAPAPAE 307
QY 301 APAPAPAGEVAPTPTTPTPQRTPLA 325
Db 308 APAPAPAGEVAPTPTTPTPQRTPLA 332

RESULT 8
AAY39083
ID AAY39083 standard; protein; 332 AA.
XX AC AAY39083;
XX DT 05-NOV-1999 (first entry)
XX DE M. tuberculosis antigen DPEP amino acid sequence.
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
XX KW immune response; skin test.
XX OS Mycobacterium tuberculosis.
XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US003268.
XX PR 18-FEB-1998; 98US-00025197.
XX PR 05-MAY-1998; 98US-00072967.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX DR WPI; 1999-527409/44.
XX DR N-PSDB; AAZ19300.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX tests and protective or therapeutic vaccines or compositions.
XX Example 1; Page 93; 299pp; English.
CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
CC polypeptides fragments, can be used in pharmaceutical compositions or
CC vaccines to generate a protective or therapeutic immune response to M.
CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
CC Ag can induce proliferation of, or cytokine secretion by, T. B or natural
CC killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
CC to AAZ19460 and AAY39083 to AAY39225 are used in the exemplification of
CC the present invention
XX SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 2; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e-104;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDNLTERRKGRLLAALAIAAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 60
 DB 8 MHQVDNLTERRKGRLLAALAIAAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 67
 QY 61 PPAPATVAPPPPPAAANTPNAQGDENAPPPADPNAPPPVIAAPNAPQVRIIDNPVGGF 120
 DB 68 PPAPATVAPPPPPAAANTPNAQGDENAPPPADPNAPPPVIAAPNAPQVRIIDNPVGGF 127
 QY 121 SFALPAGWVSDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIVLGRDQKLYASAEA 180
 DB 128 SFALPAGWVSDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIVLGRDQKLYASAEA 187
 QY 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 240
 DB 188 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 247
 QY 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEIRPLVAPPAPAPAPAPAP 300
 DB 248 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEIRPLVAPPAPAPAPAPAP 307
 QY 301 APAPAPAGEVAPTPTTPTPQRTPLA 325
 DB 308 APAPAPAGEVAPTPTTPTPQRTPLA 332

RESULT 9
 AAU01895
 ID AAU01895 standard; protein; 332 AA.
 XX AAU01895;
 AC AAU01895;
 DT 29-AUG-2001 (first entry)
 DE M. tuberculosis DPEP antigen.
 KW DPEP; antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease.
 OS Mycobacterium tuberculosis.
 XX WO200124820-A1.
 XX 12-APR-2001.
 XX 10-OCT-2000; 2000WO-US028095.
 XX 07-OCT-1999; 99US-0158338P.
 XX 07-OCT-1999; 99US-0158425P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
 XX WPI; 2001-290576/30.
 XX N-PSDB; AAS03786.
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 XX proteins comprising combinations of heterologous antigens.
 XX Disclosure; Page 160; 168pp; English.

The sequence represents Mycobacterium tuberculosis DPEP, an M.
 tuberculosis antigen. Compositions comprising at least 2 heterologous
 antigens, as a fusion protein, and vectors expressing the fusion proteins
 are used as vaccines to prophylactically immunise mammals (especially
 humans) against infection by Mycobacteria. The compositions contain at
 least 2 heterologous antigens that increase the serological sensitivity
 of individuals infected with tuberculosis, a disease frequently affecting

CC patients with acquired immunodeficiency disease, AIDS
 XX SQ Sequence 332 AA;
 Query Match 100.0%; Score 1726; DB 4; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.3e-104;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDNLTERRKGRLLAALAIAAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 60
 DB 8 MHQVDNLTERRKGRLLAALAIAAMASASLVTVAVPATANADPPAPPVPTTAASPPSTAAA 67
 QY 61 PPAPATVAPPPPPAAANTPNAQGDENAPPPADPNAPPPVIAAPNAPQVRIIDNPVGGF 120
 DB 68 PPAPATVAPPPPPAAANTPNAQGDENAPPPADPNAPPPVIAAPNAPQVRIIDNPVGGF 127
 QY 121 SFALPAGWVSDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIVLGRDQKLYASAEA 180
 DB 128 SFALPAGWVSDAAHFDYGSALLSKTTGDPFFGQPPPVANDTRIVLGRDQKLYASAEA 187
 QY 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 240
 DB 188 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPSKNGQIWT 247
 QY 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEIRPLVAPPAPAPAPAPAP 300
 DB 248 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEIRPLVAPPAPAPAPAPAP 307
 QY 301 APAPAPAGEVAPTPTTPTPQRTPLA 325
 DB 308 APAPAPAGEVAPTPTTPTPQRTPLA 332

RESULT 10
 AAE29720
 ID AAE29720 standard; protein; 332 AA.
 XX AAE29720;
 AC AAE29720;
 DT 27-JAN-2003 (first entry)
 DE Mycobacterium tuberculosis DPEP antigenic protein.
 KW Vaccine; immunity; diagnostic agent; gene therapy; DPEP antigen.
 XX Mycobacterium tuberculosis.
 XX WO200272792-A2.
 XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 XX 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Brannon M, Guderian J;
 XX WPI; 2002-759844/82.
 XX N-PSDB; AAD47097.
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeifF,
 XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 XX tuberculosis.
 XX Disclosure; Page 111; 155pp; English.

The invention relates to a recombinant nucleic acid molecule encoding a
 fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 polynucleotide sequence encoding an antigen or an antigenic fragment from
 Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a

CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
CC are useful as vaccines for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is M.
CC tuberculosis DPEP antigenic protein
XX
SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 5; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.3e-104;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDPNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
Db 8 MHQVDPNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 67
QY 61 PPAPATPVAPPPAAANTPNAQPGDFNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
Db 68 PPAPATPVAPPPAAANTPNAQPGDFNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPPEPGPPVANDTRIVLGRDOKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPPEPGPPVANDTRIVLGRDOKLYASAEA 187
QY 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPKNGQIWT 240
Db 188 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPKNGQIWT 247
QY 241 GVIGSPAANADPAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPPAPAPAPAP 300
Db 248 GVIGSPAANADPAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPPAPAPAPAP 307
QY 301 APAPAPAGEVAPTPTTPTPTPTLPA 325
Db 308 APAPAPAGEVAPTPTTPTPTPTLPA 332

RESULT 11
AAE17584
ID AAE17584 standard; protein; 332 AA.
XX
AC AAE17584;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species DPEP protein.
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; DPEP protein.
XX
OS Mycobacterium sp.
XX
FN WO200198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US019959.
XX
PR 20-JUN-2000; 2000US-00597796.
PR 01-FEB-2001; 2001US-0265737P.
XX
FA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Alderson M;
XX
DR WPI; 2002-147798/19.
XX
PT N-PSDB; AAD28355.
XX
Composition comprising MTB39 antigen and MTB32A antigen from

PT Mycobacterium species, useful for eliciting immune response in a subject.
XX
PS Claim 9; Page 127; 136pp; English.
XX
CC The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human.
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC DPEP protein
XX
SQ Sequence 332 AA;

Query Match 100.0%; Score 1726; DB 5; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.3e-104;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDPNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 60
Db 8 MHQVDPNLTTRKGRLLAALAIAAMASASLVTVAVPATANADPEAPVPVTTAASPSTAAA 67
QY 61 PPAPATPVAPPPAAANTPNAQPGDFNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
Db 68 PPAPATPVAPPPAAANTPNAQPGDFNAAAPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPPEPGPPVANDTRIVLGRDOKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPPEPGPPVANDTRIVLGRDOKLYASAEA 187
QY 181 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPKNGQIWT 240
Db 188 TDSKAAARLGSDMGDFYMPYPTGTRINQETVSLDANGVSGSASYEYVKFSDPKNGQIWT 247
QY 241 GVIGSPAANADPAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPPAPAPAPAP 300
Db 248 GVIGSPAANADPAGPPQRFVWVLTANNPVDKGAALAESIRPLVAPPPAPAPAPAP 307
QY 301 APAPAPAGEVAPTPTTPTPTPTLPA 325
Db 308 APAPAPAGEVAPTPTTPTPTPTLPA 332

RESULT 12
AAE17584
ID AAE17584 standard; protein; 325 AA.
XX
AC AAE17584;
XX
DT 03-JAN-2001 (first entry)
XX
DE Fibronectin attachment protein FAP-B.
XX
KW Inflammation; inflammatory response; irritant; pathogen; treatment;
KW T helper cell; lymphocyte; cell mediated immunity; skin allergy; hives;
KW allergic rhinitis; conjunctivitis; hay fever; allergic gastroenteritis;
KW asthma; bronchopulmonary aspergillosis; pollutant;
KW respiratory tract infection.
XX

OS Mycobacterium bovis.
 FN WO200048622-A2.
 XX 24-AUG-2000.
 PD 22-FEB-2000; 2000WO-US004463.
 XX 22-FEB-1999; 99US-0121177P.
 XX (IOWA) UNIV IOWA RES FOUND.
 PA Ratliff TL, Kline JN;
 XX WPI; 2000-549237/50.
 DR N-PSDB; AAA53822.
 XX Inhibiting an inflammatory response in a mammal for treating skin
 PT allergy, allergic rhinitis, hay fever, or asthma comprises administering
 PT a microbial polypeptide.
 PS Claim 4; Fig 1; 52pp; English.
 XX A new method of treating an inflammatory response comprises administering
 CC a microbial polypeptide which is preferably a mycobacterial polypeptide.
 CC The method is particularly useful for treating or inhibiting a Th2
 CC mediated inflammatory response. The inflammatory response is associated
 CC with a disease such as skin allergy, hives, allergic rhinitis,
 CC conjunctivitis, hay fever, allergic gastroenteritis, or asthma.
 CC Specifically, the asthma is intrinsic, i.e. associated with an irritant
 CC (e.g. a pathogen causing a respiratory tract infection in a mammal, or an
 CC inhaled pollutant). The asthma may also be extrinsic, which includes
 CC allergic asthma, occupational asthma or allergic bronchopulmonary
 CC aspergillosis
 XX
 SQ Sequence 325 AA;
 Query Match 99.7%; Score 1720; DB 3; Length 325;
 Best Local Similarity 99.7%; Pred. No. 3.2e-104;
 Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHQVDNLTTRKGLAALATAAASASLVTVAVPATANADPPAPVPVTTAASPDSTAA 60
 DB 1 MHQVDNLTTRKGLAALATAAASASLVTVAVPATANADPPAPVPVTTAASPDSTAA 60
 QY 61 PPAPATVAPPPAAAANTPNAQPGDPAAPPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
 DB 61 PPAPATVAPPPAAAANTPNAQPGDPAAPPPADPNAPPPVIAPNAPQPVRIIDNPVGGF 120
 QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGQPPVANDTRIVLGLDOKLYASAEA 180
 DB 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGQPPVANDTRIVLGLDOKLYASAEA 180
 QY 181 TDSKAAARLGSMDGEFTYMPYPTGTRINQETVSLDANGVSGSASYEYKESDPSKPNQIWT 240
 DB 181 TDSKAAARLGSMDGEFTYMPYPTGTRINQETVSLDANGVSGSASYEYKESDPSKPNQIWT 240
 QY 241 GVIGSPANAPDAGPPQRFVWVGTTANNPVDKGAALAESIRPLVAPPAPAPAPAP 300
 DB 241 GVIGSPANAPDAGPPQRFVWVGTTANNPVDKGAALAESIRPLVAPPAPAPAPAP 300
 QY 301 APAPAPAGEVAPTPTTPTQRTLPA 325
 DB 301 APAPAPAGEVAPTPTTPTQRTLPA 325
 RESULT 13
 AA003566
 ID AA003566 standard; protein; 286 AA.
 XX
 AC AA003566;
 XX
 DT 22-APR-1997 (first entry)

XX Mycobacterium tuberculosis antigenic determinant mature protein.
 DS
 XX Antigenic determinant; Mycobacterium tuberculosis; ion exchange; human;
 XX chromatography; gel filtration; reverse phase column chromatography;
 KW immunogenic; serum; guinea pig; expression vector; cosmid; antibody;
 KW Mycobacterium smegmatis; Mycobacterium bovis BCG; microorganism; vaccine;
 XX hybrid; epitope; disease; diphtheria; cholera; toxin.
 OS Mycobacterium tuberculosis.
 XX
 XX WO9623885-A1.
 XX 08-AUG-1996.
 XX 31-JAN-1996; 96WO-FR000166.
 XX 01-FEB-1995; 95US-00382184.
 XX (INSP) INST PASTEUR.
 XX Laqueyrie A, Marchal G, Pescher P, Romain F;
 XX WPI; 1996-371433/37.
 XX N-PSDB; AAT39357.
 XX Mycobacterium tuberculosis antigen and hybrid proteins comprising it -
 PT useful in vaccines against tuberculosis and in immunoassays.
 XX
 PS Claim 1; Page 49; 74pp; French.
 CC This is the amino acid sequence of the mature portion of an antigenic
 CC determinant protein from Mycobacterium tuberculosis. The mature protein
 CC has calculated mol. wt. of 28779 Da but has an experimental mol. wt. of 45
 CC -47 kD as determined by SDS-PAGE. The difference is thought to be due to
 CC the high frequency of Pro residues (21.7%) in the sequence. The protein
 CC was purified from M. tuberculosis strain H37Rv by conventional
 CC chromatographic methods e.g. low pressure ion exchange and reverse phase column
 CC 300 gel filtration, DEAE ion exchange and SDS-PAGE and
 CC chromatography. The proteins were then separated by SDS-PAGE and
 CC immunogenic proteins determined with immune serum from humans with TB or
 CC guinea pigs infected with M. tuberculosis. A complex of mol. wt. 45-47 kD
 CC was identified. An expression library of genomic M. tuberculosis DNA was
 CC generated in the cosmid vector pTUS18. The library was transformed into
 CC M. smegmatis. Clones expressing antigenic determinants were isolated by
 CC screening with an antibody against the M. bovis BCG 45/47 kD proteins. 3
 CC clones were isolated, all of which contained identical sequences i.e. the
 CC sequence presented here. The protein or microorganisms expressing it can
 CC be used as vaccines against tuberculosis. Also hybrid proteins comprising
 CC this protein and epitopes from other disease causing organisms or
 CC proteins, e.g. diphtheria or cholera toxin, can be used as vaccines
 CC against their respective diseases
 XX
 SQ Sequence 286 AA;
 Query Match 89.3%; Score 1542; DB 2; Length 286;
 Best Local Similarity 100.0%; Pred. No. 1.1e-92;
 Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 40 DEEPAPVPTTAAASPDSTAAAPATVAPPPAAAANTPNAQPGDPAAPPPADPNAPP 99
 DB 1 DEEPAPVPTTAAASPDSTAAAPATVAPPPAAAANTPNAQPGDPAAPPPADPNAPP 60
 QY 100 PPVIAPNAPQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPDPGQPPV 159
 DB 61 PPVIAPNAPQPVRIIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPDPGQPPV 120
 QY 160 ANDTRIVLGRDOKLYASAEATDSKAAARLGSMDGEFTYMPYPTGTRINQETVSLDANGVSG 219
 DB 121 ANDTRIVLGRDOKLYASAEATDSKAAARLGSMDGEFTYMPYPTGTRINQETVSLDANGVSG 180
 QY 220 SASYVEKFSDPSPKNGQIWTGVIKSPANAPDAGPPQRFVWVGTTANNPVDKGAALAESIRPLVAPPAPAPAPAP 279

CTC FIVE LETTERS FOR A TWO LETTER WORD

Qy	227	KFSDPSKNGQIWTGVIGSPAANAPDAGPPQRFVWVWLTANNPVDKGAAKALAESIRPL	286
Db	554	KFSDPSKNGQIWTGVIGSPAANAPDAGPPQRFVWVWLTANNPVDKGAAKALAESIRPL	613
Qy	287	VAPPPAPAPAPAPAPAPAGEVAPTPTTPTQRTLPA	325
Db	614	VAPPPAPAPAPAPAPAPAGEVAPTPTTPTQRTLPA	652

Search completed: July 7, 2004, 18:21:03
 Job time : 50.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:19:15 ; Search time 22.5 Seconds
(without alignments)
1389.433 Million cell updates/sec

Title: US-10-720-192-3
Perfect score: 1726
Sequence: 1 MHQVDPNLTRKGRLLAALAI.....PAGEVAPPTPTPTPQRTLPA 325

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1726	100.0	325	2 D70666	probable modD prot
2	980	56.8	287	2 B87166	hypothetical prote
3	241	14.0	1585	2 T31611	hypothetical prote
4	237.5	13.8	1188	2 S49915	extensin-like prot
5	228.5	13.2	464	2 S22637	extensin - Volvox
6	226.5	13.1	222	2 H96711	hypothetical prote
7	224.5	13.0	3534	2 T42587	tegument protein 2
8	219.5	12.7	85	2 T44744	antigen 43L [impor
9	217.5	12.6	817	2 S51342	verprolin - yeast
10	217.5	12.6	839	2 T75518	hypothetical prote
11	217	12.6	801	2 T29018	hypothetical prote
12	216.5	12.5	847	2 P96531	hypothetical prote
13	213.5	12.4	760	2 T06291	extensin homolog 1
14	213.5	12.4	3164	1 WNEB56	Uu36 protein - hum
15	208.5	12.1	214	2 T10737	extensin-like cell
16	207	12.0	416	1 SKXL4G	dermal gland prote
17	206.5	12.0	1611	2 T38236	hypothetical prote
18	204.5	11.8	539	2 T28770	hypothetical prote
19	203	11.8	1201	2 G86441	unknown protein [1
20	203	11.8	3421	1 W2BE56	36/K tegument prot
21	202.5	11.7	283	2 S13383	hydroxyproline-ric
22	201.5	11.7	395	2 H75457	hypothetical prote
23	200.5	11.6	214	2 T09854	proline-rich cell
24	200.5	11.6	760	2 F86387	probable Pro kinas
25	199	11.5	708	2 D96711	hypothetical prote
26	199	11.5	839	2 T04859	extensin homolog F
27	197.5	11.4	865	2 A47282	calcium-binding pr
28	196.5	11.4	418	2 T15442	hypothetical prote
29	196.5	11.4	666	2 B70803	hypothetical prote

30	194	11.2	191	2 F84522	probable proline-r
31	193.5	11.2	1151	2 T18535	high molecular mas
32	193	11.2	620	2 S06733	hydroxyproline-ric
33	192.5	11.2	731	2 T04455	hypothetical prote
34	192.5	11.2	744	2 E86255	hypothetical prote
35	192.5	11.2	873	2 A47283	calphotin fruit
36	191.5	11.1	280	2 T11671	extensin-like prot
37	191	11.1	576	2 T36729	probable serine/th
38	190	11.0	550	2 G70597	probable proteinas
39	189.5	11.0	267	2 S08314	cell wall glyccopro
40	189	11.0	240	2 B24264	proline-rich prote
41	189	11.0	1794	2 T38459	hypothetical diver
42	188	10.9	544	2 T17547	proline-rich prote
43	188	10.9	699	2 C43674	US4 protein - huma
44	188	10.9	1288	2 T31420	C-terminal domain-
45	187	10.8	350	2 S22456	hydroxyproline-ric

ALIGNMENTS

RESULT 1

D70666

probable modD protein - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: D70666

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devaiah, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;

Rajandream, M.A.; Rogers, J.; Rubner, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulstom, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70666

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-325 <COL>

A:Cross-references: GB:283859; GB:AL123456; NID:g3261678; PIDN:CAB06127.1; PID:e290722;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: modD

Query Match 100.0%; Score 1726; DB 2; Length 325;

Best Local Similarity 100.0%; Pred. No. 2.8e-87;

Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDNLTTRKGRLLAALAIAMASSLVTVAVPATANADPEPAPVPTTAASPFSTAAA 60

Db 1 MHQVDNLTTRKGRLLAALAIAMASSLVTVAVPATANADPEPAPVPTTAASPFSTAAA 60

Qy 61 PPAPATPVAPPFAAANTPNAPQGDNPAPPPADPNAPPPVIAPNAPQVRIIDNPVGGF 120

Db 61 PPAPATPVAPPFAAANTPNAPQGDNPAPPPADPNAPPPVIAPNAPQVRIIDNPVGGF 120

Qy 121 SFALPAGWESDAAHFDYGSALLSKTTGDPDPGPPPPVANDTRIVLGLDQKLYASAEA 180

Db 121 SFALPAGWESDAAHFDYGSALLSKTTGDPDPGPPPPVANDTRIVLGLDQKLYASAEA 180

Qy 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKESDPKNGQIWT 240

Db 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKESDPKNGQIWT 240

Qy 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKESDPKNGQIWT 240

Db 181 TDSKAAARLGSDMGFFYMPYPTGTRINQETVSLDANGVSGSASYEVKESDPKNGQIWT 240

Qy 241 GVIGSPAANAPDAGPPQRFWVVLGTANNPVDKGAALKALAESIRPLVAPPAPAPAPAE 300

Db 241 GVIGSPAANAPDAGPPQRFWVVLGTANNPVDKGAALKALAESIRPLVAPPAPAPAPAE 300

Qy 301 APAPAPAGEVAPPTPTPTPQRTLPA 325

Db 301 APAPAPAGEVAPPTPTPTPQRTLPA 325

RESULT 2

B87166
 hypothetical protein modD [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: B87166
 R: Cole, S.T.; Biglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt,
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A86909; PMID:21128732; PMID:11234002
 A:Accession: B87166
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-287 <STO>
 A:Cross-references: GB:AL450380; NID:gl3093663; PIDN:CAC31010.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: modD

Query Match 56.8%; Score 980; DB 2; Length 287;
 Best Local Similarity 66.8%; Pred. No. 8.8e-47;
 Matches 199; Conservative 21; Mismatches 64; Indels 14; Gaps 3;

QY 1 MHQVDNLTTRKRLAALAAASASLVTVAVPATANADPEAPVPTTAASPPSTAAA 60
 DB 1 MNQVLDLSTHRKGLWAILAIAVVASAFTMLPAAANADPAFL-----PPSTATA 51

QY 61 PPAPATVAPPAAANTPNACQDPPNAPPPADNAPPVPIAPNAPQPVRIQNPVGGF 120
 DB 52 APSAQEITTPUGAPVSEKQDPPNAP--PSLDNAPYPLAVDPNAG--RITNAVGGF 106

QY 121 SPALPAGVSDAAHFDYGSALLSKTTGDPDPGPPQPPVANDTRIVLGRDQKLYASAEA 180
 DB 107 SFVLPAWVSEASHLDYGSVLLSKAEQPPVGLQPTVAVTDTRIVLGRDQKLYASAEA 166

QY 181 TSKAAALRGDSMGFFYPGPTNRQETVSLDANGVSGSASYEVKFSKDPKNGQIWT 240
 DB 167 DNKAARVLGDSMGFFYLPPTNRQETVSLDANGVSGSASYEVKFSKDPKNGQIWT 226

QY 241 GVIGSPANAPDAGPPQRFVWVLTANNPDKGAALAESIRPLVAPPAPAPA 298
 DB 227 SVVGSFASSTPDVGSQRFVWVLTGTSNNPDKGAALAESIRSEMAPIASVSAPA 284

RESULT 3
 T31611
 hypothetical protein Y50B8A.g - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T31611
 R:Steward, C.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: Z21047
 A:Accession: T31611
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-1385 <WIL>
 A:Cross-references: EMBL:AL117200; NID:el549770; PIDN:CAB55050.1; CESP:Y50B8A.g
 A:Experimental source: clone Y50B8A
 C:Genetics:
 A:Gene: CESP:Y50B8A.g
 A:introns: 25/3; 60/1; 133/2; 217/3; 270/3; 337/2; 400/1; 746/2

Query Match 14.0%; Score 241; DB 2; Length 1585;
 Best Local Similarity 26.8%; Pred. No. 7.1e-06;
 Matches 97; Conservative 12; Mismatches 122; Indels 134; Gaps 13;

QY 41 PFAPPPVPTTAASPPS-----TAAAPPAPATVAPPAPAA 75
 DB 1233 PPPPPAPAPAPAPSGGSGGSAAGGGGGGGYTGSSAAPPFPFPFPFPFP 1292

QY 76 ANTPNAQPG-----DPNAAPPADPNAPPPVPIAPNPQVR 112

DB 1293 APAPAPSSGGYSGSGSAAAGGGGGSGGYSGGSAAPPPPPPPPPAPAP-APAP-S 1350
 QY 113 IDNPVGGSFALPAGWVESDAAHFDYGSALLSKTTGDPDPGPPQPPVANDTRIVLGRDQ 172
 DB 1351 SGYSGSGGSAAGGGGGSGGYTGSSA---APPFPFPFPFPFPFPAP----- 1395
 QY 173 KLYASAEATDSKAAARLGSMDGMEFTNYPYPTNRQETVSLDANGVSGSASYEVKFSKDP 232
 DB 1396 ---APAPAPSSGGYSG-GSSGGS-----AAGGGGGSGGYSGGSAAPP 1434
 QY 233 KNGQIWTGICSPANAPDAGPPQRFVWVLTANNPDKGAAXA-----LAESIRPLVA 288
 DB 1435 PP-----PPAPAPAPAPSGGYSG--GSSGSAAGGGGGSGGYTGSSAAP--- 1478
 QY 289 PPPAP 320
 DB 1479 PPPPPPPPPPPPPAP 1538
 QY 321 RTLPA 325
 DB 1539 PPPPA 1543

RESULT 4
 S49915
 extensin-like protein - maize
 C:Species: Zea mays (maize)
 C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
 C:Accession: S49915
 R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
 submitted to the EMBL Data Library, June 1994
 A:Description: Pex genes: pollen-specific genes with extensin-like domains.
 A:Reference number: S49915
 A:Accession: S49915
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1188 <RUB>
 A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 13.8%; Score 237.5; DB 2; Length 1188;
 Best Local Similarity 28.9%; Pred. No. 8.2e-06;
 Matches 87; Conservative 30; Mismatches 123; Indels 61; Gaps 13;

QY 34 PATANADPEP--APPVPTTAASPPSTAAAPPAPAPATVAPPVAPAAANTPNACPDNAP 91
 DB 569 PAPVASPPFPVKSPPPPTLVASPPFPVKSPPPA-PVASPPFPVKSP--PPTPVASPP 624

QY 92 PADPNA-PPPPVITAPNAPQVRIDNPVGGSFALPAGWVESDAAHFDYGSALLSKTTGDP 150
 DB 625 PPAPVASPPFPVKSP--PPTPVSSPPFPVKSPPPPPPPPPAKSTPPPEEYPTPTSVKSSP 682

QY 151 PFPQPPPPVANDTRIVLGRDQKLYASAEATDSKAAARLGSMDGMEFTNYPYPTNRQETV 210
 DB 683 PEKSLPPP-----TLISPPPPQEK--FTPTSTPSKPS-----SPEKP 718

QY 211 SLDANGVSGSASYEVKFSKDPKNGQIWTGICSPANAPDAGPPQRFVWVLTANNP 270
 DB 719 SPKPEPVSSPPQ--TPKSPPPAPVSPSPPTPVSSPPALAPVSSPP-----SVKSSP 768

QY 271 VDKGAALAESIRPLVAPPPAP-----APAPAB---PAPAPAPAGVAPPTTPTPTQL 323
 DB 769 PP-----APLSPPPAPQVKSPPFPVQVSSFPFPAPKSSPLAPVSSPPQVEKTS 817

QY 324 P 324
 DB 818 P 818

RESULT 5
 S22697
 extensin - Volvox carteri (fragment)
 C:Species: Volvox carteri

Db	28	AAGPVTSTTTAPPTTAAAPTAAAPPTTTTTPVSAQAQPPASPVTTPPAVTTPTSPAPKV	87
QY	76	AN--TPNAQPGDNAAAPPADENAPPPVATPNAPQVRI DNVGGFSFALPAGWVESDA	133
Db	88	AFVISAPATPPQPPQSPAGAPT VSPPVSPFPAP	122
QY	134	AHFDYGSALLSKTTGGPPFPFGQPPPVANDTRI VLGRLDQKLYASAEATDSKAAARLGSDM	193
Db	123	-----TSPPPTTASPPP	144
QY	194	GEFMPYPGTRI NETVSLDANGVSGSASYEVKFS DPSPKPNQGIWTCVIGSFAANAPDA	253
Db	145	-----SPPAP	161
QY	254	GPPQRFVYVLGTANNPV DKGAALAESIRPLVAPPAPAPAPAE	300
Db	162	-----ISJPPAPAPATPKHKRKHKKHHHAP	198
QY	301	APAPAPAGEVAPPTTPTTPTORTLPA	325
Db	189	APAPIPPSPSP-FVLITDQDTAPA	212
RESULT 7			
T42567			
Segment protein 24 - equine herpesvirus 4 (strain NS80567)			
C:Species: equine herpesvirus 4			
A:Variety: strain NS80567			
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000			
C:Accession: T42567			
R:Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.			
J. Gen. Virol. 79, 1197-1203, 1998			
A:Title: The DNA sequence of equine herpesvirus-4.			
A:Reference number: Z22173; MUID:98264457; PMID:960335			
A:Accession: T42567			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-3534 <TEL>			
A:Cross-references: EMBL:AF030027; MID:g2605950; PIDN:AACS9539.1; PID:g2605967			
A:Experimental source: strain NS80567			
C:Genetics:			
A:Note: 24			
C:Superfamily: varicella-zoster virus gene 22 protein			
Query Match			
Best Local Similarity 13.0%; Score 224.5; DB 2; Length 3534;			
Matches 91; Conservative 18; Mismatches 93; Indels 125; Gaps 13;			
QY	5	DNPLTRKGRGLAALAIAMASASLVTVAVPATANADPEAPVPVTTAASP--PSTAAAPP	62
Db	2710	DPN---EALLTAPSKPAAPAPAPSKPAAPAPSKPAAPAPSKPAAPAPSKPAAPAPSKPAAPAPSK	2765
QY	63	APATPVAPPPPPAAANTENAPQGDNPAAAPPADPNAPPPPVIAFNAPQVRI DNVGGFSF	122
Db	2766	KFAAAPAPSKPAAPAPSK-----KPAAPAPSKPAAPAPSKPAAPAPSK-----	2811
QY	123	ALPAGWVESDAAHFDYGSALLSKTTGCDPFPQCPPPVANDTRI VLGRLDQKLYASAEATD	182
Db	2812	--PAA-----APAPSKP	2828
QY	183	SKAAARLGSDMGCEFFYMPYPTGTRI NETVSLDANGVSGSASYEVKFS DPSPKPNQGIWTCV	242
Db	2829	KFAAA-----PAPSKP-----	2839
QY	243	IGSPAAANAPDAGF-PQRVFWVLGTANNPV DKGAALAESIRPLVAP---PPAPAPAPA	298
Db	2840	AAAPAPSKFAAPAPSK-----PAAAPAPSKFAAPAPSK-KPAAAPAPSKFAAPAPSK	2891
QY	299	EPAPAPAPAGEVAPPTTPTTPTORTLPA	325
Db	2892	KFAAAPAPSK-KFAAAPAPSKPQNTLVA	2917

RESULT 8

T44744
antigen 43L [imported] - Mycobacterium leprae (fragment)
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T44744
R:Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z22833
A:Accession: T44744
A:Status: preliminary; translated from GB/EMBL/DDSV
A:Molecule type: DNA
A:Residues: 1-85 <PAR>
A:Cross-references: EMBL:AL008609; PIDN:CAA15433.1
A:Experimental source: cosmid B1788
C:Genetics:
A:Note: 43L

Query Match 12.7%; Score 219.5; DB 2; Length 85;
Best Local Similarity 52.1%; Pred. No. 5.7e-06;
Matches 50; Conservative 8; Mismatches 27; Indels 11; Gaps 2;
QY 1 MHQVDPNLTTRKGLAALATAMASASLVTVAVPATANADPEPAPPVPTTAASPFSTAAA 60
DB 1 MNQVLDSTTRKGLWALATAVASASFTMPLPAAANADPAPL-----PPSTATA 51
QY 61 PPAPATVAPPPPPAAANTPNAQPDNAPPPADPN 96
DB 52 APSPAQEIITPLGAPVSVSEAPGPDNA--PSLDPN 85

RESULT 9

S51342
verpulin - yeast (Saccharomyces cerevisiae)
N:Alternate names: prolin-rich protein VRP1; protein L8300.13; protein YLR337C
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1995 #sequence_revision 11-Aug-1995 #text_change 23-Mar-2001
C:Accession: S51342; S39626; S57435
R:Du, Z.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8300.
A:Reference number: S51339
A:Accession: S51342
A:Molecule type: DNA
A:Residues: 1-817 <DUZ>
A:Cross-references: EMBL:U19028; NID:G609380; PID:G609382; MIPS:YLR337C
R:Donnelly, S.F.H.; Pocklington, M.J.; Pallotta, D.; Orr, E.
Mol. Microbiol. 10, 585-596, 1993
A:Title: A proline-rich protein, verpulin, involved in cytoskeletal organization and cell division in yeast.
A:Reference number: S39626; MUID:95058201; PMID:7968536
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-307, 'R', 309-349, 'R', 351-688, 'E', 690-709, 'HLRWIPVPLIAPVKTLNNGYFLOVDRRCNTSII
A:Cross-references: EMBL:Z26645; NID:G414785; PIDN:CAA81388.1; PID:G439289
R:Munn, A.L.; Stevenson, B.J.; Gali, M.I.; Riezman, H.
submitted to the EMBL Data Library, June 1995
A:Description: end5, end6, and end7: mutations that cause actin delocalization and block cell growth.
A:Reference number: S57435
A:Accession: S57435
A:Molecule type: DNA
A:Residues: 1-162, 'F', 164-817 <MUN>
A:Cross-references: EMBL:X87806; NID:G871534; PID:G871535
A:Experimental source: strain W303
C:Genetics:
A:Gene: SGD:VRP1; MDP5; END5
A:Cross-references: SGD:S0004329; MIPS:YLR337C
A:Map position: 12R
F:5-14/Region: proline-rich
F:77-85/Region: proline-rich
F:114-180/Region: proline-rich
F:216-245/Region: proline-rich
F:305-336/Region: proline-rich

F:349-357/Region: proline-rich
F:372-382/Region: proline-rich
F:396-406/Region: proline-rich
F:421-445/Region: proline-rich
F:518-528/Region: proline-rich
F:567-577/Region: proline-rich
F:608-621/Region: proline-rich
F:649-661/Region: proline-rich
F:678-685/Region: proline-rich
F:704-710/Region: proline-rich
Query Match 12.6%; Score 217.5; DB 2; Length 817;
Best Local Similarity 25.3%; Pred. No. 6.9e-05;
Matches 99; Conservative 31; Mismatches 125; Indels 131; Gaps 18;
QY 24 ASASLVTVAVPATA-----NADPEPAPPVPT--TAASPP-----STAAAPPAPATEVAPP 72
DB 118 ASAPPPIGAVPSVAAPPIENAPLSPAPAVPSIPSSSAPPIDIPSSAAPPPIVPSPPAP 177
QY 73 P-----AAANTENAPQPDNAP-----PPA-----DPN 96
DB 178 PLPLSGASAPKVPQNRPMPSVRPAHRSHQRKSSNISLPSVSAPPLPSASLPTHVSPPQ 237
QY 97 APPPPV-----IAP--NAPOVVRIDNPVGGFSPALPAGWVESDAAHFDYGS----- 140
DB 238 APPPPPTTIGLDSKNIKPTDNAVSPSPSEVPAGGLPEL-----AEINARSRSEGAEGV 292
QY 141 -----ALLSKTTGDPFPFGQPPPVAND-----TRVLGRLOOKLYASAEATDSKAAA 187
DB 293 SSTKIOTENHKSPQFPFLFSSAPPITPTSHAPPLPTAPPPLPNTVTSAPKATSAFAPP 352
QY 188 RLGSMDGEFMPYPTGRINQETVSLDANGVSGSASVYEVKFSKNGQITWGVIGSPA 247
DB 353 P-----PELPAAMSSASTNSVKATVPPTLA-----PPLPN-----TTSVPENKA 392
QY 248 ANAPDAGPPQRFVWVLTGANNPVDKGA---AKALAESIRPLVAPPAPAPAPAPAPAPAP-A 303
DB 393 SGNPAPPPPP-----PPPGAFSTSSALSASSIPLAPLPPPPPPSVATSPESA 440
QY 304 PAPAGEVAPTPTTPTP 319
DB 441 PPP-----PPTLTNKP 452
RESULT 10
F75518
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75518
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F.
M.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma,
S.; Shen, M.; Matheson, J.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75518
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <WHI>
A:Cross-references: GB:AE001904; GB:AE000513; NID:G6458129; PIDN:AAF10038.1; PID:G645814;
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0458
A:Map position: 1
C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
Query Match 12.6%; Score 217.5; DB 2; Length 839;
Best Local Similarity 25.4%; Pred. No. 7.1e-05;
Matches 95; Conservative 24; Mismatches 124; Indels 131; Gaps 17;
QY 22 AVASASLVTVAVPATANAD-----PEPAP--PVPTTAASPFSTAAAP-----PAPATPVAP- 70

Db 91 APATKPTPTAKPATPAPEPPKPPPTTPEPKPETPEPPKPAPEPPKPDTPAEPLKP 150
Qy 71 -----PPAAANTPNAGQDNPNAAPPADENAPP-----PPVIANV-----PQ 109
Db 151 VODTPPPVTPKVTPEPVTPEKAPTPEPVLOPVAQTTPVAKPPVPAPTSTPTPPVQPA 210
Qy 110 PVRID-----NPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDP-----PPPGQPPP 158
Db 211 PRTTPPPQAARTPNAPATQTPAPAQAQAPATQTPATPAPAAQRPAAGAPSP 270
Qy 159 VANDTRIVLRDQKLYASAEATDSKAAARLGSMDGEFYMYPGTRI-----NQE 208
Db 271 -----APAPA-----QANAPAGSVVPEATVPESSTPAAPSQAQTPPTRE 310
Qy 209 TVSLDANGVSGSASYEYKFSDESKPNQGIWTGVIGSPAANAPAGPQRFVFWLGTAN 268
Db 311 TAQTEAS-----PAAPNS-----SAAAPNEPASEP-----VAG 338
Qy 269 NPVDKGAKALAEIRPLVAPP-----PAPAFAPAPAPAPAPAPAPAPAPAPAP 305
Db 339 RP---GTAASGPESASPVTTPRGETPDTAASAGTPSAGRVTPAPAPASASEGSAARTPG 395
Qy 306 PAGEVAPPTTPPTP 319
Db 396 AGSQTPPIPATPIP 409

RESULT 11

T29018

hypothetical protein ZK84.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C:Accession: T29018

R:Kirsten, J.

submitted to the EMBL Data Library, April 1995

A:Description: The sequence of C. elegans cosmid ZK84.

A:Reference number: 220553

A:Accession: T29018

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-801 <KIR>

A:Cross-references: EMBL:U23181; PIDN:AAC48204.1; GSPDB:GN00020; CESP:ZK84.1

A:Experimental source: strain Bristol N2; clone ZK84

C:Genetics:

A:Gene: CESP:ZK84.1

A:Map position: 2

A:Insertions: 22/2; 45/3; 108/1

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 12.6%; Score 217; DB 2; Length 801;
Best Local Similarity 24.9%; Pred. No. 7.2e-05;
Matches 101; Conservative 37; Mismatches 151; Indels 116; Gaps 16;

Qy 25 SASLVTVAVPATANADPEPAP-----FVPTTAAS---PPST----- 57
Db 194 ASGVTTSEQEGAGDAFERAPVVEETPAPTAAEETAPATEASNAVGTPEGYVDGTANA 253
Qy 58 --AAAPAPATPVAPPAAANTP-NAQPDGNAPPPADPNAPPVPPVIANPAPQPVRI 114
Db 254 ASAAEAPAVETPA--PAPAAEETAPATSESEAPAPAPATAETPAPETVSA-APEAANSY 310
Qy 115 NPVGGFSFALPAGWVES-----DAAHFDYGSALLSKTTGD----- 149
Db 311 DSAGGDAATAPSSSEADAAFTDSASADTTAALVDTSSSEHAESTEAPATDIANTE 370
Qy 150 -PPFGGPPPVAN-----DTRIVLRDQKLYASAEATDSKAAARLGSMDGEF 196
Db 371 TTPAPSVAPVADAAAGYDSPTSIPETTPAPAAEDTPAPASAAAEETAPAPAAE--- 426
Qy 197 YMPVPGTR-----INQETVSLDANGVSGSASYEYKFSDESKPNQGI 238
Db 427 -TPAPETASAPDAGGAAPADVAAPADVATTAPETSSAQSAAGSYDV-----PSEPA 481

Qy 239 WTGVIGS-----PAANADAGDPQ--RMFVVVLGTANNPVDK--GAAKALAE 281
Db 482 TAPIVESATEAPSDSAAPIGFAASEPAPAPATDAATLETAPAPAAEPAPAPAAAG 541
Qy 282 SIRPLVAPPAPAPAP-ABPAPAPAPAGVAPTPTTTPQRTPLPA 325
Db 542 YDAPSSVPETTPAPAPADETPAPAPAAETTPAPAPAAETPAPA 586

RESULT 12

P96531

hypothetical protein F13F21.7 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: P96531

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: P96531

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-847 <STO>

A:Cross-references: GB:AE005173; NID:G5430752; PIDN:AAD43152.1; GSPDB:GN00141

C:Genetics:

A:Gene: F13F21.7

A:Map position: 1

Query Match 12.5%; Score 216.5; DB 2; Length 847;
Best Local Similarity 26.0%; Pred. No. 8.1e-05;
Matches 79; Conservative 35; Mismatches 135; Indels 55; Gaps 12;

Qy 33 VPATANADPEPAPPVTTAASPSTAAAPAPATPVAP---PPAAANTPNAGQDNPNA 88
Db 533 VFPQPPMSPSPSPPIYPPPPVHSPPPPVYSSPPPHVYSPPPVVASPPPPPPV 592

Qy 89 APPADPNAPPVPIAPNAPQFVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTG 148
Db 593 SPPEPVPFSPPEVTSPPPPSPVYSPPPP---SHSPP-----PPVVS 631

Qy 149 DPFPQCPQPPVANDTRIVLG--RLQKLYASAEATDSKAAARLGSMDGEFYMYPDG-TRI 205
Db 632 PPPTFTSPPTHTNQPPMGAPTPTCAPTSPSETTQVTPSS-ESDQQLSPVQAPTPV 690

Qy 206 NOETVSLDANGV---SGSASYEYKFSDFSKPNQGIWTGVIGSPAANAPADGPPQRFVWV 262
Db 691 QSSTPSSETPQVPTPSSSEYQAPNLSPVQAPT-----PVQAPTTSETSOVP----- 738

Qy 263 WLGTTANNPVDKAAKALAEIRPLVAPP-----APAPAPAPAPAPAPAGV-APTPT 315
Db 739 ---TPSSSESNQSQAPTPILEPVHAPTNSKPVQSPTPS-SEPVSSPEQSEVEAPEPT 794

Qy 316 TPTP 319
Db 795 PVNP 798

RESULT 13

T06291

extensin homolog T9E8.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 11-Jun-1999

C:Accession: T06291

R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schuell

submitted to the Protein Sequence Database, April 1999

A:Reference number: Z15588

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:17:39 ; Search time 10.5 Seconds
(without alignments)
1611.695 Million cell updates/sec

Title: US-10-720-192-3
Perfect score: 1726
Sequence: 1 MHQVDPNLTNRKRLAALAI.....PAGEVAPTTTPTQRTILPA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	1726	100.0	325	1 APA_MYCTU	Q50906 mycobacteri
2	1720	99.7	325	1 APA_MYCBO	Q30620 mycobacteri
3	985	57.1	381	1 APA_MYCAV	Q48919 mycobacteri
4	980	56.8	287	1 APA_MYCLE	P46842 mycobacteri
5	243	14.1	555	1 G1P1_CHLRE	Q9fpg6 chlamydomon
6	217.5	12.6	817	1 VRP1_YEAST	P37370 saccharomyc
7	213.5	12.4	3164	1 TEGU_HSV11	P10220 herpes simp
8	210	12.2	1083	1 T2D3_HUMAN	Q00268 homo sapien
9	207	12.0	439	1 XP2_YENLA	P37437 xenopus lae
10	205.5	11.9	802	1 ENAH_MOUSE	Q31173 mus musculu
11	203	11.8	3421	1 TSGU_HSV2B	P28955 equine herp
12	197.5	11.4	865	1 CPN_DROME	Q02910 drosophila
13	197	11.4	283	1 EXTN_SORBI	P24152 sorghum bic
14	193	11.2	620	1 EXTN_TOBAC	P13983 nicotiana t
15	192.5	11.2	1198	1 HCN4_RAT	Q9jka7 rattus norv
16	191	11.1	1386	1 ZAP3_MOUSE	Q9r017 mus musculu
17	189.5	11.0	267	1 EXTN_MAIZE	P14918 zea mays (m
18	189	11.0	1794	1 YDC9_SCHPO	Q10172 schizosacch
19	189	11.0	2167	1 SHK1_RAT	Q9wv48 rattus norv
20	188.5	10.9	2715	1 MLL4_HUMAN	Q9umh6 homo sapien
21	188	10.9	699	1 VGLG_HSV2H	P13290 herpes simp
22	186.5	10.8	5179	1 MUC2_HUMAN	Q02817 homo sapien
23	186	10.8	331	1 PRP1_HUMAN	P04280 homo sapien
24	186	10.8	1033	1 IF2_STRCO	Q8cjq8 streptomyc
25	185.5	10.7	1300	1 SAL3_HUMAN	Q9bxa3 homo sapien
26	184.5	10.7	639	1 ZIC5_HUMAN	Q96t25 homo sapien
27	184.5	10.7	1046	1 IP2_STRAW	Q82k53 streptomyc
28	184	10.7	1271	1 RBMG_HUMAN	Q9upn6 homo sapien
29	183	10.6	347	1 CSP_PLABA	P23033 plasmodium
30	182	10.5	261	1 PRP2_MOUSE	P05142 mus musculu
31	180.5	10.5	534	1 APG_ARATH	P40602 arabidopsis
32	180.5	10.5	2716	1 OSA_DROME	Q8in94 drosophila
33	179.5	10.4	2142	1 BAT2_HUMAN	P48634 homo sapien

ALIGNMENTS

RESULT 1

ID	APA_MYCTU	STANDARD	PRT	325 AA.
AC	Q50906; Q08062;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Alanine and proline-rich secreted protein apa precursor (45/47 kDa			
DE	antigen) (Fibronectin attachment protein) (immunogenic protein MPT32)			
DE	(Antigen MPT-32) (45-kDa glycoprotein) (FAP-B).			
GN	APA OR MODD OR RV1860 OR MT-908 OR MTCY359.13.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RC	MEDLINE=96009758; PubMed=7558311;			
RA	Laqueyrie A., Maltizer P., Romain F., Biglmeier K., Cole S.,			
RA	Marchel G.			
RA	"Cloning, sequencing, and expression of the apa gene coding for the			
RT	Mycobacterium tuberculosis 45/47-kilodalton secreted antigen			
RT	complex."			
RL	Infect. Immun. 63:4003-4010(1995).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RA	Laqueyrie A.			
RA	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RL	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RC	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			

open

RL J. Bacteriol. 184:5479-5490(2002).
RN [5]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=9109989; PubMed=1898999;
RA Nagai S., Wiker H.G., Harboe M., Kinomoto M.;
RT "Isolation and partial characterization of major protein antigens in
RL the culture fluid of Mycobacterium tuberculosis";
RL Infect. Immun. 59:372-382(1991).
RN [6]
RP PARTIAL SEQUENCE, AND GLYCOSYLATION.
RX MEDLINE=95347792; PubMed=7622204;
RA Dobos K.M., Swiderek K., Khoo K.-H., Brennan P.J., Belisle J.T.;
RT "Evidence for glycosylation sites on the 45-kilodalton glycoprotein of
RL Mycobacterium tuberculosis";
RL Infect. Immun. 63:2846-2853(1995).
RN [7]
RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=96196153; PubMed=8626314;
RA Dobos K.M., Khoo K.-H., Swiderek K.M., Brennan P.J., Belisle J.T.;
RT "Definition of the full extent of glycosylation of the 45-kilodalton
RL glycoprotein of Mycobacterium tuberculosis";
RL J. Bacteriol. 178:2498-2506(1996).
RN [8]
RP SEQUENCE OF 40-49, AND CHARACTERIZATION OF CARBOHYDRATE-LINKAGE SITES
BY MASS SPECTROMETRY.
RX STRAIN=H37Rv;
RX MEDLINE=20011399; PubMed=10542234;
RA Horn C., Namane A., Pescher P., Riviere M., Romain F., Puzo G.,
Barzu O., Marchal G.;
RT "Decreased capacity of recombinant 45/47-kDa molecules (Apa) of
RL Mycobacterium tuberculosis to stimulate T lymphocyte responses
related to changes in their mannosylation pattern";
RL J. Biol. Chem. 274:32023-32030(1999).
RN [9]
RP SEQUENCE OF 40-57, AND CHARACTERIZATION OF CARBOHYDRATE-LINKAGE SITES
BY MASS SPECTROMETRY.
RX STRAIN=H37Rv;
RX MEDLINE=20002527; PubMed=10531201;
RA Romain F., Horn C., Pescher P., Namane A., Riviere M., Puzo G.,
Barzu O., Marchal G.;
RT "Deglycosylation of the 45/47-kilodalton antigen complex of
RL Mycobacterium tuberculosis decreases its capacity to elicit in vivo
or in vitro cellular immune responses";
RL Infect. Immun. 67:5567-5572(1999).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MASS SPECTROMETRY: MW=28782; MW_ERR=1.6; METHOD=Electrospray;
CC RANGE=40-325.
CC -1- BIOTECHNOLOGY: Major immunodominant antigen that has potential as
CC a vaccine against tuberculosis. APA-ELISA could be used in
CC diagnosis.
CC -1- MISCELLANEOUS: Changes in the mannosylation pattern of this
CC protein affect its ability to stimulate T-lymphocyte response.
CC -1- CAUTION: Was originally thought to be involved in molybdenum
CC transport.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X80268; CAA56555.1; --
CC EMBL; X99258; CAA67645.1; --
CC EMBL; Z83859; CAB06127.1; --
CC EMBL; AE007048; AAK46179.1; --
CC PIR; D70666; D70666.
CC TIGR; MT1908; --
CC TubercuList; Rv1860; --
KW Antigen; Glycoprotein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 339
FT CHAIN 40 325 ALANINE AND PROLINE-RICH SECRETED PROTEIN

FT DOMAIN 85 107 APA.
FT REPEAT 85 88 N-A.
FT REPEAT 94 97 1.
FT REPEAT 104 107 2.
FT CARBOHYD 49 49 O-LINKED (MAN. . .).
FT CARBOHYD 57 57 O-LINKED (MAN. . .).
FT CARBOHYD 66 66 O-LINKED (MAN. . .).
FT CARBOHYD 316 316 O-LINKED (MAN. . .).
FT CONFLICT 136 136 F -> L (IN REF. 4).
SQ SEQUENCE 325 AA; 32720 MW; 59E5D045A997BED CRC64;

Query Match 100.0%; Score 1726; DB 1; Length 325;
Best Local Similarity 100.0%; Pred. No. 1.5e-80;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDNLTERRKRLAALAAIAAASASLVTVAVPATANADPEAPVPTTAASPSSTAAA 60
DB 1 MHQVDNLTERRKRLAALAAIAAASASLVTVAVPATANADPEAPVPTTAASPSSTAAA 60
QY 61 PPAPATFPVAPPPAAANTNAQPDNAAPPADPNAPPPVIAAPVIAAPVIAAPVIAAPV 120
DB 61 PPAPATFPVAPPPAAANTNAQPDNAAPPADPNAPPPVIAAPVIAAPVIAAPVIAAPV 120
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFFGQPPVANDTRIVLGRDQKLYASAEA 180
DB 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFFGQPPVANDTRIVLGRDQKLYASAEA 180
QY 181 TDSKAAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYEYKFSKPKNGQIWT 240
DB 181 TDSKAAARLGSDMGFEYMPYPTGTRINQETVSLDANGVSGSASYEYKFSKPKNGQIWT 240
QY 241 GVIGSPAANAPDAGPPQRFVWLGTANNPVDKGAALAEISIRPLVAPPPAPAPAPAP 300
DB 241 GVIGSPAANAPDAGPPQRFVWLGTANNPVDKGAALAEISIRPLVAPPPAPAPAPAP 300
QY 301 APAPAPAGEVAPT 325
DB 301 APAPAPAGEVAPT 325

RESULT 2
APA_MYCBO STANDARD; PRT; 325 AA.
AC O30620;
DI 30-MAY-2000 (Rel. 39, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DI 15-MAR-2004 (Rel. 43, Last annotation update)
DE Alanine and proline-rich secreted protein apa precursor (Fibronectin
DE attachment protein) (45/47 kDa antigen) (FAP-B).
GN APA OR MB1891.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BCG;
RA Zhao W., Schorey J.S., Bong-Mastek M., Brown E.J., Ratliff T.L.;
RT "Identification, sequence and characterization of the M. bovis BCG
RT fibronectin attachment protein";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=AF2122/97; PubMed=12788972;
RX MEDLINE=22709107;
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Dushoff S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.F., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).

```

CC      -!- SUBCELLULAR LOCATION: Secreted.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF013569; AAB71842.1; -.
CC      EMBL; BX248340; CAD94594.1; -.
CC      XN Antigen; Glycoprotein; Repeat; Signal; Complete proteome.
CC      SIGNAL 1 39 BY SIMILARITY.
CC      CHAIN 40 325 ALANINE AND PROLINE-RICH SECRETED PROTEIN
CC      FT DOMAIN 85 107 3 X 4 AA APPROXIMATE REPEATS OF [DA]-P-
CC      FT REPEAT 85 88 1. N-A.
CC      FT REPEAT 94 97 2.
CC      FT REPEAT 104 107 3.
CC      FT CARBOHYD 49 49 O-LINKED (MAN. .) (BY SIMILARITY).
CC      FT CARBOHYD 57 57 O-LINKED (MAN. .) (BY SIMILARITY).
CC      FT CARBOHYD 66 66 O-LINKED (MAN. .) (BY SIMILARITY).
CC      FT CARBOHYD 316 316 O-LINKED (MAN. .) (BY SIMILARITY).
CC      SQ SEQUENCE 325 AA; 32686 MW; D3419CA5547D91E9 CRC64;
CC
CC      Query Match 99.7%; Score 1720; DB 1; Length 325;
CC      Best Local Similarity 99.7%; Pred. No. 3e-80;
CC      Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC
QY      1 MHQVDNLTNRKGRALALAIAMASASLVTVAVPATANADPEPPVPTTAASPPSTAAA 60
DB      1 MHQVDNLTNRKGRALALAIAMASASLVTVAVPATANADPEPPVPTTAASPPSTAAA 60
QY      61 PPAPATPVAPPPPPAAANTPNAPQGDPAAPPPADPNAPPPVIAFNAPQPVRIENPVGGF 120
DB      61 PPAPATPVAPPPPPAAANTPNAPQGDPAAPPPADPNAPPPVIAFNAPQPVRIENPVGGF 120
QY      121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGPPQPPVANDTRIVLGRDLQKLYASAEA 180
DB      121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFGPPQPPVANDTRIVLGRDLQKLYASAEA 180
QY      181 TDSKAAARLGSMDGFEYMPYFPTRIHQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240
DB      181 TDSKAAARLGSMDGFEYMPYFPTRIHQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240
QY      241 GVIGSPAANADAGPPQRFVWVLTANNPVVDKGAALAESIRPLVAPPPAPAPAPAP 300
DB      241 GVIGSPAANADAGPPQRFVWVLTANNPVVDKGAALAESIRPLVAPPPAPAPAPAP 300
QY      301 APAPAGEVAPTPTTPTPTPTPTLPA 325
DB      301 APAPAGEVAPTPTTPTPTPTPTLPA 325
CC
CC      RESULT 3
CC      APA_MYCAV
CC      ID APA_MYCAV STANDARD; PRT; 381 AA.
CC      AC Q49519.
CC      DT 30-MAY-2000 (Rel. 39, Created)
CC      DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC      DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC      DE Alanine and proline-rich secreted protein apa precursor (45/47 kDa
CC      antigen) (Fibronectin attachment protein) (FAP-A).
CC      GN APA.
CC      OS Mycobacterium avium.
CC      OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC      OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC      OX NCBI_TaxID=1764;
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RX STRAIN=101;

```

```

RX      MEDLINE=97011577; PubMed=8858587;
RA      Schorey J.S., Holsti M.A., Ratliff T.L., Allen P.M., Brown E.J.;
RT      "Characterization of the fibronectin-attachment protein of
RT      Mycobacterium avium reveals a fibronectin-binding motif conserved
RT      among mycobacteria.";
RL      Mol. Microbiol. 21:321-329(1996).
CC      -!- SUBCELLULAR LOCATION: Secreted.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U53585; AAB50543.1; -.
CC      XN Antigen; Signal.
CC      SIGNAL 1 32 POTENTIAL.
CC      CHAIN 33 381 ALANINE AND PROLINE-RICH SECRETED
CC      FT PROTEIN APA.
CC      SQ SEQUENCE 381 AA; 38210 MW; ODACB9A22AA11D2B CRC64;
CC
CC      Query Match 57.1%; Score 985; DB 1; Length 381;
CC      Best Local Similarity 55.2%; Pred. No. 2.7e-43;
CC      Matches 218; Conservative 27; Mismatches 66; Indels 84; Gaps 12;
CC
QY      1 MHQVDNLTNRKGRALALAIAMASASLVTVAVPATANADPEPPVPTTAASPP----- 55
DB      1 MDQVEATSTRKGLWTLTATTVSGASAVVIALPETSHADPEVPTVPPESTATTTPRRR 60
QY      56 -----STAAAPPAPA-----TPVAPPAAANTPNAPQCDNAA-PPPA 93
DB      61 RIFGQADNAQAGAPAPANGQQRPRRRMPTTRAPPAGAPPPGAPPAAPNGAPPPV 120
QY      94 DPNAPPPPVIAFNAPQPVRIENPVGGFSAFALPAGWVESDAAHFDYGSALLSKTTGDPFPP 153
DB      121 DPNAPPPPADPNAG--RIEN-----SYVLPAWVESDASHLDYGSALLSKVTGPPMP 172
QY      154 GQPPPVANDTRIVLGRDLQKLYASAEATSKAAARLGSMDGFEYMPYPTRIHQETVSLD 213
DB      173 DQPPPVANDTRIVMGSDVQKLYASAEANNAKAAVGLGSDMGFEYMPYPTRIHQETVSLD 232
QY      214 -ANGVSGSASYEVKFSDFSKPNQIWTGVIGSPAANADAGPPQRFVWVLTANNPV 272
DB      233 GANGSTGSASYEVKFSDFSKPNQIWTGVIGS-----ANAGNQRQRFVWVLTANNPV 287
QY      273 KGAALKALAESIR---PIVAPPPAP-----APAPAP- 300
DB      288 KVAALKALAESIQAWTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 347
QY      301 -----APAP-----APAGEVAPTPTTPTPTPTLPA 325
DB      348 APAAPGAPAPGQAPAVEVSPPT-TPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 381
CC
CC      RESULT 4
CC      APA_MYCLE
CC      ID APA_MYCLE STANDARD; PRT; 287 AA.
CC      AC P46842; O32905; Q9RSV6;
CC      DT 01-NOV-1995 (Rel. 32, Created)
CC      DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC      DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC      DE Alanine and proline-rich secreted protein apa precursor (Antigen 43L)
CC      (Fibronectin attachment protein) (FAP-L).
CC      GN APA OR WDDO OR ML2055 OR MLCE1788.C1C.
CC      OS Mycobacterium leprae.
CC      OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC      OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CC      OX NCBI_TaxID=1769;
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RX MEDLINE=94086110; PubMed=8262636;

```

RA Wiele B., van Agterveld M., Janson A., Clark-Curtiss J.E.,
 RA Rinke de Wit T.F., Harboe M., Thole J.,
 RT "Characterization of a Mycobacterium leprae antigen related to the
 RT secreted Mycobacterium tuberculosis protein MPT32.",
 RL Infect. Immun. 62:252-258(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95310024; PubMed=7790081;
 RA Schorey J.S., Li Q., McCourt D.W., Bong-Mastek M., Clark-Curtiss J.E.,
 RA Ratliff T.L., Brown E.J.,
 RA "A Mycobacterium leprae gene encoding a fibronectin binding protein is
 RT used for efficient invasion of epithelial cells and Schwann cells.",
 RL Infect. Immun. 63:2652-2657(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.,
 RA "Massive gene decay in the leprosy bacillus.",
 RL Nature 409:1007-1011(2001).
 RL Nature 409:1007-1011(2001).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X76501; -; NOT ANNOTATED_CDS.
 DR EMBL; AL583924; CAC31010.1; -;
 DR EMBL; AL008609; CAA15433.1; -;
 DR PIR; B87166; B87166.
 DR Leptoma; M2055; -;
 KW Antigen; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 39
 FT CHAIN 40 287 ALANINE AND PROLINE-RICH SECRETED PROTEIN
 FT
 FT DOMAIN 76 96 3 X 4 AA REPEATS OF D-P-N-A.
 FT REPEAT 76 96 1.
 FT REPEAT 83 86 2.
 FT REPEAT 93 96 3.
 FT CONFLICT 33 34 LP -> FR (IN REF. 1).
 SQ SEQUENCE 287 AA; 29559 MW; 7FC860A8B3026CFB CRC64;
 Query Match 56.8%; Score 980; DB 1; Length 287;
 Best Local Similarity 66.8%; Pred. No. 3.7e-43;
 Matches 199; Conservative 21; Mismatches 64; Indels 14; Gaps 3;
 QY 1 MHQVDNLTTRGRLLAALAAWASASLVTVAVPATANADPEAPVPTTAASPESTAAA 60
 DB 1 MNQVDLSTHRLGLWAILAIAVVASASAPTEPLPAAANADAPL-----PPSTATA 51
 QY 61 PPAPATPVAPPFAAANTNAQGDPAAPPADNAPPPVPIAPNAPQPVRIQNPVGGF 120
 DB 52 AFSQAQEIITPLPGAPVSSEAQGDPA--PSLDENAPVPLAVDPAAG---RIITNAVGGF 106
 QY 121 SFALPAGWESAAAHFDYGSALLSKTTGDPDPFPVPPGPPVANDTRIVLGRDQKLYASAE 160
 DB 107 SFLVLPAGWVESAAHLDDYGSALLSKAIBQPPVILGQPTVVATDTRIVLGRDQKLYASAE 166
 QY 181 TDSKAAARLGSMDGFFMPYFGPTRINQETVSLDANGVGSASYYEVKFSDFSKNGQIWT 240
 DB 167 DNKAARVLGSDMGFFLYFPQTRINQETIPLHANGIAGSASYYEVKFSDFSKNGQICT 226

QY 241 GVTGSPAANAPDAGPPQRFVVLGTANNPVDKGAALAEISIRLIVAPPPAPAPA 298
 DB 227 SVVGSAASTPDVGPQRWFWVLGTNNPVDKGAALAEISIRSEMAPPASVSAPA 284

RESULT 5
 GPI_CHLRE STANDARD; PRT; 555 AA.
 AC Q9FP06; Q03927;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
 DE glycoprotein 1).
 GN GPI.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 OX NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21159032; PubMed=11258910;
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
 RA Goodenough U.W.,
 RA "Glycosylated polyproline II rods-with-kinks as a structural motif in
 RT plant hydroxyproline-rich glycoproteins.",
 RL Biochemistry 40:2978-2987(2001).
 RN [2]
 RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=91017504; PubMed=1699225;
 RA Adair W.S., Apt K.E.;
 RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
 RT encoding cell wall hydroxyproline-rich glycoproteins.",
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
 CC -!- FUNCTION: Major component of the outer cell wall W6 (crystalline)
 CC layer.
 CC -!- SUBUNIT: Associates with GP2 and GP3.
 CC -!- PTM: N-glycosylated and O-glycosylated.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF309494; AAG45420.1; -;
 DR EMBL; MS8496; AAG69706.1; ALT_SEQ.
 DR GlycoSuiteDB; Q9FP06; -;
 DR InterPro; IPR003882; Pistil extensin.
 DR PRINTS; PR01218; PSTLEXTENSIN.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
 FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSX REPEATS.
 FT DOMAIN 259 279 POLY-PRO.
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 555 AA; 54219 MW; 6A584A9045502F5 CRC64;
 Query Match 14.1%; Score 243; DB 1; Length 555;
 Best Local Similarity 24.6%; Pred. No. 6.8e-06;
 Matches 70; Conservative 21; Mismatches 69; Indels 124; Gaps 8;

QY 39 ADPEPPVPTTAAASPPSTAAAPAPATPV--APPFAAANTNAQGDPAAPPADN 96
 DB 203 APPSPAPPVPPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 262
 QY 97 APPPPVPIAPNAPQPVRIQNPVGGFSPALPAGWVESDAAHFDYGSALLSKTTGDP 156


```
Db 263 APPPPPPPPPPPP-----RP-----PPFA-----NTPMPSPSPSP 293
QY 157 PPVANDTRIVLGRDLQKLYASAEATDSKAAARLGSMDGMEFYMPGTRINQETVSLDANG 216
Db 294 PP-----SPAPPT-----301
QY 217 VSGSASYEVKFSKNGKNGIQTWGVISPAANAPDAGPPQRFVWVWLTANNPVDKGA 276
Db 302 -----PTPPSPSPSPSPVPSPAPVPSPSPSPSP-----329
QY 277 KALAESIRPLVAPPAPAPAPAPAPAPAPAGEVAPTPT-TPTP 319
Db 330 -----PAPSPSPSPAPPTSPSPSPSPSPSPSPSPSPSPSPSPSP 365
```

RESULT 6

```
VRP1 YEAST
ID VRP1 YEAST STANDARD; PRT; 817 AA.
AC P37370; Q06133;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Verprolin.
GN VRP1 OR MDP2 OR END5 OR YLR337W OR L8300.13.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A364;
RX MEDLINE=95058201; PubMed=7968536;
RA Donnelly S.F.H., Pocklington M.J., Pallota D., Orr E.;
RT "A proline-rich protein, verprolin, involved in cytoskeletal
RT organization and cellular growth in the yeast Saccharomycetes
RT cerevisiae.";
RL Mol. Microbiol. 10:585-596(1993).
RN [2]
```

SEQUENCE FROM N.A.

```
RC STRAIN=S288c / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston W., Hillier L., Riles L., Albertmann K., Andre B., Anserge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Dueterhoeft A.,
RA Enkian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Netwlich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetalle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomycetes cerevisiae chromosome XII.";
RL Nature 387:87-90(1997)
```

CC -!- FUNCTION: Involved in cytoskeletal organization and cellular growth. May exert its effects on the cytoskeleton directly, or indirectly via proline-binding proteins (e.g. profilin) or proteins possessing SH3 domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; Z26645; CAB81388.1; -
CC DR EMBL; U19028; AAB67263.1; -
CC DR PIR; S51342; S51342.
CC DR GeneOnline; 142401; -
CC DR SGD; S0004329; VRP1.
CC GO; GO:0005857; C:actin cortical patch (sensu Saccharomycetes); IDA.
```

```
DR GO; GO:0003779; F:actin binding; IDA.
DR GO; GO:0007015; P:actin filament organization; IPI.
DR GO; GO:0006997; P:endocytosis; IMP.
DR GO; GO:0007121; P:polar budding; IMP.
DR GO; GO:0006370; P:response to osmotic stress; IMP.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 2.
DR SMART; SM00246; WH2; 2.
KW Cytoskeleton; Repeat.
FT DOMAIN 5 14 POLY-PRO.
FT DOMAIN 239 245 POLY-PRO.
FT DOMAIN 349 357 POLY-PRO.
FT DOMAIN 396 406 POLY-PRO.
FT DOMAIN 424 431 POLY-PRO.
FT DOMAIN 462 468 POLY-SER.
FT DOMAIN 704 708 POLY-PRO.
FT CONFLICT 308 308 P -> R (IN REF. 1).
FT CONFLICT 350 350 A -> R (IN REF. 1).
FT CONFLICT 689 689 V -> E (IN REF. 1).
FT CONFLICT 710 817 PSTMDTGTSSPSKLNKQRLFGSTGLQHKHNTHTNQPDV
DVGRTYIGGSSIVGAKSGNERIVIDDSRQKWNVQMPKP
RPFQNTKLYPSGKSSVFLDLFT -> HLRWIPVPLIA
PVKTLNNGYFLQVDRRCNTSIIRIQNQMLM (IN REF.
1).
SQ SEQUENCE 817 AA; 82593 MW; 24C752D5B1CA1C8 CRC64;
```

Query Match 12.6%; Score 217.5; DB 1; Length 817;

Best Local Similarity 26.3%; Pred. No. 0.00018;

Matches 99; Conservative 31; Mismatches 125; Indels 121; Gaps 18;

QY 24 ASASLVTVAVPATA-----NADPEAPPVPT--TAASPP-----STAAAPPAPATPVAPP 72

Db 118 ASAPPITGAVPSVAAPPINAPLSPAPAVPSIPSSSAPPIPIPIVPSAP 177

QY 73 P-----AAANTNAOFGDPNAAP-----PPA-----DPN 96

Db 178 FLPLSGASAPKVPQNRPHMPSVRPAHRSHQKSSNISLPSVSAPPLPSASLPTHVSNPPQ 237

QY 97 APPPPV-----IAP--NAPQVRIDNPVGGFSEALPAGWSDAAHPDYG-----140

Db 238 APPPTPTTIGLDSKNIKPTDNAVSPSPSEVPAGGLPFL-----ABINARRSERGAVEG 292

QY 141 -----ALLSKTTGDPFFGQPPPVAND-----TRIVLGRDLQKLYASAEATDSKAAA 187

Db 293 SSTKIQTENHKSPSQPPLPSSAPPIPTSHAPPLPTAPPPLPNVTSAPKKATSAPAP 352

QY 188 RLQSDGFEFMPYPGTRINQETVSLDANGVSGSASYEVKFSKNGKNGIQTWGVISPA 247

Db 353 P-----PPLPAAMSSASTNSVKATVPPTLA-----PPLFN---TTSVPFNKA 392

QY 248 ANAPDAGPPQRFVWVWLTANNPVDKGA---AKALAESIRPLVAPPAPAPAPAEAP-A 303

Db 393 SSNPAPPPPP-----PPPGAFTSSALSASSIPLAPLPPPPPSVATSVPSA 440

QY 304 PAPAGEVAPTPTPTPT 319

Db 441 PPP-----PPTLTNKP 452

RESULT 7

```
TEGU_HSV11
ID TEGU_HSV11 STANDARD; PRT; 3164 AA.
AC P10220;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Large tegument protein (Varion protein UL36).
GN UL36.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
```



```

RP SEQUENCE FROM N.A.
RX MEDLINE=68274327; PubMed=2839594;
RA McGeoch D.J., Dallymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
RT herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574 (1988).
CC -!- FUNCTION: Tegument protein.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X14112; CAA32311.1; -
CC PIR; I30085; WMBE6.
CC InterPro; IPR006928; Herpes teg N.
CC InterPro; IPR005210; Herpes teg N.
CC Pfam; PF04843; Herpes teg N.
CC Pfam; PF03586; Herpes_UL36; 1.
KW Repeat.
FT DOMAIN 2911 3980 35 X 2 AA TANDEM REPEATS OF P-Q.
SQ SEQUENCE 3164 AA; 335857 MW; CCSD31FF4F9FE3F4 CRC64;
Query Match 12.4%; Score 213.5; DB 1; Length 3164;
Best Local Similarity 26.5%; Pred. No. 0.001;
Matches 91; Conservative 17; Mismatches 117; Indels 119; Gaps 15;
QY 10 RKGRALALAIAMASA-----SLTVAVPATANADPEPAPVPTAASPSTAAAP 61
Db 2683 RRRHRRSLARATQASATCGWRPPLDPTVAVPTDFARP-PAPKPPPEPAPHALVSGVP 2741
QY 62 -----PAPATPVAP-PPPAANTNAGDPNAPPAAPPAAPPPVADNAPQP 110
Db 2742 LPLGPOAQAQSPALDIDVPPVATGT--VLPGCENRRPPLTSPATPRPVGGQR 2799
QY 111 VRIDNPVGGFSPALPAGWVESDAHFYGSALLSKTGDPPPGPPPVANDTRVLGLRL 170
Db 2800 -RLTRP-----AVASLSSESRLSPFMDP---ADPTAPVLGR- 2832
QY 171 DQKLYASAEATSKAARLGDGMEFYMPYPCPTRINQETVSLDANGVSGSASYEVEKFS 230
Db 2833 -----NPAPETSSSPAG----- 2844
QY 231 PSKPNQIWTGIVGSPAANAPDAGPPQRFVFWVLGTA-----NNPVDKGAALAESIR 284
Db 2845 PSEPPPAV-----QVAPPTSGPPTVLTLEGGVAPGVPVSRRTTRQPVATTTTSAR 2898
QY 285 P-----LVAPPAPAPAPAPAPAPAPAGEVAPTPTTTPQ 320
Db 2899 PRGHLTVSLSAPQPPQPPQPP-QPQPPQPPQPPQPP-QPQPPQ 2940
RESULT 8
T2D3_HUMAN
ID T2D3_HUMAN STANDARD; PRT; 1083 AA.
AC 000268; Q99721; Q9BR40; Q9BX42;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcription initiation factor TFIIID 135 kDa subunit (TAFII-135)
DE (TAFII135) (TAFII-130) (TAFII130)
GN TAF4 OR TAF4A OR TAF2C1 OR TAF2C OR TAFII135 OR TAFII130.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=97336072; PubMed=9192867;
RA Mengus G., May M., Carre L., Chambon P., Davidson I.;
RA "Human TAF(II)135 potentiates transcriptional activation by the AF-2s
RT of the retinoic acid, vitamin D3, and thyroid hormone receptors in
RT mammalian cells.";
RL Genes Dev. 11:1381-1395 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley C.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.F., Clee C.M.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.F., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivaeslaih M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLeay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Suce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RL "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871 (2001).
RN [3]
RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97098442; PubMed=8942982;
RA Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIID
RT complex: hTAFII130 and hTAFII100.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616 (1996).
RN [4]
RP IDENTIFICATION IN THE TFIID-HAT COMPLEX WITH TAF5L; TAF6L; TADA3L;
RX SUPT3H; TAF2; TAF5; TTRAP; GCN5L2 AND TAF10.
RX MEDLINE=99303588; PubMed=10373431;
RA Brand M., Yamamoto K., Staub A., Tora L.;
RT "Identification of TAF4-binding protein-free TAFII-containing complex
RT subunits suggests a role in nucleosome acetylation and signal
RT transduction.";
RL J. Biol. Chem. 274:18285-18289 (1999).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 870-918 IN COMPLEX WITH
RX TAF12.
RX MEDLINE=20063193; PubMed=10594036;
RA Gangloff Y.-G., Werten S., Romier C., Carre L., Poch O., Moras D.,
RA Davidson I.;
RT "The human TFIID components TAF(II)135 and TAF(II)20 and the yeast
RT SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like
RT pairs";
RL Mol. Cell. Biol. 20:340-351 (2000).
CC -!- FUNCTION: Makes part of TFIID is a multimeric protein complex that
CC plays a central role in mediating promoter responses to various
CC activators and repressors. Potentiates transcriptional activation
CC by the AF-2s of the retinoic acid, vitamin D3 and thyroid hormone.
CC -!- SUBUNIT: TFIID is composed of TATA binding protein (TBP) and a
CC number of TBP-associated factors (TAFs). Component of the TFIID-HAT
CC complex, at least composed of TAF5L, TAF6L, TADA3L, SUPT3H/SPT3,
CC TAF2/TAFII50, TAF4/TAFII135, TAF5/TAFII100, GCN5L2/GCN5, TAF10
CC and TTRAP.
CC -!- SUBCELLULAR LOCATION: Nuclear.

```

```

CC  -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC  -!- SIMILARITY: Contains 1 TAFH/NHR1 domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch)
CC  -----
DR  EMBL; Y11354; CAA72189.1; -.
DR  EMBL; AL137077; CAC36006.1; -.
DR  EMBL; AL109911; CAC22312.2; -.
DR  EMBL; U75308; AAC50901.1; -.
DR  PDB; 1H3O; 26-SEP-02.
DR  TRANSFAC; T02328; -.
DR  Genew; HGNC:11537; TAF4.
DR  MIM; 601796; -.
DR  GO; GO:0005669; C:transcription factor TFIID complex; TAS.
DR  GO; GO:0016251; F:General RNA polymerase II transcription fac. .; TAS.
DR  GO; GO:0005515; F:protein binding; TAS.
DR  GO; GO:0003713; F:transcription co-activator activity; TAS.
DR  InterPro; IPR007900; TAF4.
DR  InterPro; IPR003894; TAF_hom.
DR  Pfam; PF05236; TAF4; 1.
DR  SMART; SM00549; TAFH; 1.
DR  TRANSCRIPTION REGULATION; Nuclear protein; 3D-structure.
FT  DOMAIN 590 682 TAFH/NHR1.
FT  DOMAIN 39 42 POLY-HIS.
FT  DOMAIN 52 57 POLY-ALA.
FT  DOMAIN 98 101 POLY-GLY.
FT  DOMAIN 142 148 POLY-ALA.
FT  DOMAIN 268 275 POLY-PRO.
FT  DOMAIN 331 337 POLY-ALA.
FT  DOMAIN 680 683 POLY-PRO.
FT  DOMAIN 808 813 POLY-ALA.
FT  DOMAIN 828 831 POLY-ASP.
FT  CONFLICT 105 117 PGPSRPRPLVPA -> GRGLLQRRGRES
      (IN REF. 3).
FT  CONFLICT 136 136 A -> S (IN REF. 2).
FT  CONFLICT 185 185 G -> GPG (IN REF. 2).
FT  CONFLICT 233 264 MISSING (IN REF. 3).
FT  CONFLICT 293 293 P -> L (IN REF. 3).
SQ  SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;

Query Match      12.2%; Score 210; DB 1; Length 1083;
Best Local Similarity 27.8%; Pred. No. 0.00056;
Matches 97; Conservative 20; Mismatches 140; Indels 92; Gaps 13;

QY 15 LAALAIAMASGLVTVAVPATANADPEAPVPVTTAASPSTAAAPAPATPVAPPPPA 74
DB 242 VGLAAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 300
QY 75 AAN--TPNAQCPDNAPPPADPNAPPPVVIAPNAPOF-VRIDNPVGSGFALPAGWV-- 129
DB 301 AQNGSAGAPAPAPAGAGGAGVSGQPGGAGAAAAPAPGVKAESPKRKVVQAAPEAAQT 360
QY 130 -----ESDAAHFDYG-----SALLSKTGTDPFGPPGPPVANDTRIVGLDQKLYASAEAT 181
DB 361 ASGPASTAASVMVIGTQWGLPSPAAPPAPGPTGLPKG---AGAVTQSLSRTPAT 417
QY 182 DSKAAARLGSDMGFFYMPYPGTRINQETVSLDANGVSGSASYEVKFSDFSKP-----NGQ 237
DB 418 TSGIRATLTPTVLAPRLQP-----PQNPNINQNFQ 448
QY 238 IWTGVISPAANADAGSPQRFVWLTGANNPVNDKGAALAE-STRLVADPPAPAPA 296
DB 449 LPPGQVNLVRSNGOLLMTPO-----QALAQVQAHAQAPOTTWAPR 489
QY 297 PAEPAPAP-----AP-----AGEVAPT-----PTTPTPQRTL 323
DB 490 PATTSAPPVQISTVQAPGTPIIARQVTPTTIKQVSQAQTIVQPSATL 538

```

RESULT 9

```

XP2_XENIA STANDARD; PRT; 439 AA.
ID XP2_XENIA
AC P17437; Q08944;
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Skin secretory protein XP2 precursor (APSG protein).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A. (ISOFORM 2).
RX MEDLINE=9232564; PubMed=1629230;
RA Hauser F., Roeben C., Hoifmann W.;
RT "XP2, a new member of the P-domain peptide family of potential growth
RT factors, is synthesized in Xenopus laevis skin.";
RL J. Biol. Chem. 267:14451-14455(1992).
RN [2]
RP SEQUENCE OF 3-439 FROM N.A. (ISOFORM 1).
RC TISSUE=Skin;
RX MEDLINE=90127399; PubMed=2298293;
RA Gmachl M., Berger H., Thalhammer J., Kreil G.;
RT "Dermal glands of Xenopus laevis contain a polypeptide with a highly
RT repetitive amino acid sequence.";
RL FEBS Lett. 260:145-148(1990).
CC -!- FUNCTION: May act as a growth factor in the germinal layer of the
CC epidermis. May also be involved in growth of regenerating glands
CC and in protection of the skin from the external environment.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=APEG;
CC IsoId=P17437-1; Sequence=Displayed;
CC Name=2; Synonyms=XP2;
CC IsoId=P17437-2; Sequence=VSP_004652;
CC -!- TISSUE SPECIFICITY: Skin.
CC -!- SIMILARITY: Contains 2 P-type (trefoil) domains.
CC -!- CAUTION: Ref.2 sequence differs from that shown from position 392
CC onward and is shorter (418 AA) due to a frameshift.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR  EMBL; M90095; AAA50001.1; -.
DR  EMBL; X51394; CAA35759.1; ALT_FRAME.
DR  PIR; A37331; A37331.
DR  PIR; S07498; SKXLAG.
DR  HSSP; P04155; 1PS2.
DR  InterPro; IPR000519; P_trefoil.
DR  Pfam; PF00088; trefoil; 2.
DR  PRINTS; PR00680; PTREFOIL.
DR  SMART; SM00018; PD; 2.
DR  PROSITE; PS00025; P_TREFOIL; 2.
DR  Signal; Growth factor; Alternative splicing; Repeat;
DR  Pyroliidone carboxylic acid.
DR  SIGNAL 1 22 POTENTIAL.
FT  CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
FT  MOD_RES 23 23 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).
FT  DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP] (2,4)-A-E.
FT  DOMAIN 350 391 P-TYPE 1.
FT  DOMAIN 397 438 P-TYPE 2.
FT  DISULFID 351 377 BY SIMILARITY.
FT  DISULFID 361 376 BY SIMILARITY.

```


FT	SEQUENCE	802 AA; 85844 MW; 592BB3975EE20F7F CRC64;	/FTId=VSP 007260.
SQ	Query Match	11.9%; Score 205.5; DB 1; Length 802;	
	Best Local Similarity	22.9%; Pred. No. 0.00072;	
	Matches	78; Conservative 27; Mismatches 112; Indels 123; Gaps 11;	
QY	33 VPATANADPEAPAPPVTTAAAPPSTAA	-----PPAPATPVADPP	72
DB	343 VERPLNKNRSPSSPVNTSSQPPAKSCAWTNSFSLPPSPPIMISSPFGKATGPRVL	402	
QY	73 PAAANTPNAQ-----PGDNNAA	-----PPADENAPPPVPIAPNAPOVRIDNPVG	118
DB	403 PUCVSSPVQMPSPSTAENGSLDSVTYPVSPPTSGPAAPPPPPPPPPPPPP	454	
QY	119 GFSFALPAGVSDAAHFDYGSALLSKVTGPPPPGQP-PPVANDTRIVLGRLOKLYAS	177	
DB	455 -----PPLPPPLPLAS	467	
QY	178 ABATDSKAAARLGSQWGFYMPYPCGTRI-----NQETVSLDANGVSSAS	228	
DB	468 -----LSHCGSCAPPPGCTPLASTPSSKSPSVLPSPGAPASATPLNPELGD	515	
QY	229 SDPSKPNQIWTGVTGSPAANADAGPPQRFVVLGTANNPVDKGAALAEISIRPLVA	288	
DB	516 SSASEP-----GLQAASQABSPPTQGLVLGPAPPPPPPLPSGPAVASALPPPPG	566	
QY	289 ---PPAPAPAPAPAPAPAPAGEVAPTPTTPTPORTLPA	325	
DB	567 PPPPPPLPSTGPP	606	
RESULT 11	TSQU HSVEB		
ID	TSQU HSVEB	STANDARD; PRT; 3421 AA.	
AC	P28955;		
DT	01-DEC-1992 (Rel. 24, Created)		
DT	01-DEC-1992 (Rel. 24, Last sequence update)		
DT	01-APR-1993 (Rel. 25, Last annotation update)		
DE	Large tegument protein.		
CN	24		
OS	Equine herpesvirus type 1 (strain Ab4p) (EHV-1).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Alphaherpesvirinae; Varicelloviruses.		
OX	NCBI_TaxID=31520;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
EX	MEDLINE=92295566; PubMed=1318606;		
RA	Telford E.A.R., Watson M.S., McBride K., Davison A.J.;		
RT	"The DNA sequence of equine herpesvirus-1."		
RL	Virology 189:304-316(1992).		
CC	-!- FUNCTION: Tegument protein.		
CC	-!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,		
CC	EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M86664; AA02459.1; -		
DR	PIR; G36797; WZBEE6.		
DR	InterPro; IPR006928; Herpes teg N.		
DR	InterPro; IPR005210; Herpes UL36.		
DR	Pfam; PF04843; Herpes teg N; 1.		
DR	Pfam; PF03586; Herpes_UL36; 1.		
SQ	SEQUENCE 3421 AA; 367078 MW; 5075EFB4739BB7AC CRC64;		
Query Match	11.8%; Score 203; DB 1; Length 3421;		
Best Local Similarity	23.9%; Pred. No. 0.0036;		
Matches	78; Conservative 27; Mismatches 112; Indels 123; Gaps 11;		
QY	33 VPATANADPEAPAPPVTTAAAPPSTAA	-----PPAPATPVADPP	72
DB	343 VERPLNKNRSPSSPVNTSSQPPAKSCAWTNSFSLPPSPPIMISSPFGKATGPRVL	402	
QY	73 PAAANTPNAQ-----PGDNNAA	-----PPADENAPPPVPIAPNAPOVRIDNPVG	118
DB	403 PUCVSSPVQMPSPSTAENGSLDSVTYPVSPPTSGPAAPPPPPPPPPPPPP	454	
QY	119 GFSFALPAGVSDAAHFDYGSALLSKVTGPPPPGQP-PPVANDTRIVLGRLOKLYAS	177	
DB	455 -----PPLPPPLPLAS	467	
QY	178 ABATDSKAAARLGSQWGFYMPYPCGTRI-----NQETVSLDANGVSSAS	228	
DB	468 -----LSHCGSCAPPPGCTPLASTPSSKSPSVLPSPGAPASATPLNPELGD	515	
QY	229 SDPSKPNQIWTGVTGSPAANADAGPPQRFVVLGTANNPVDKGAALAEISIRPLVA	288	
DB	516 SSASEP-----GLQAASQABSPPTQGLVLGPAPPPPPPPPLPSGPAVASALPPPPG	566	
QY	289 ---PPAPAPAPAPAPAPAPAGEVAPTPTTPTPORTLPA	325	
DB	567 PPPPPPLPSTGPP	606	
RESULT 12	CPN DROME		
ID	CPN DROME	STANDARD; PRT; 865 AA.	
AC	Q02910;		
DT	01-OCT-1993 (Rel. 27, Created)		
DT	01-OCT-1993 (Rel. 27, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Calphotin.		
CN	CPN OR CAP.		
OS	Drosophila melanogaster (fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
EX	STRAIN=Canton-S;		
RA	MEDLINE=93185729; PubMed=8094559;		
RT	Martin J.H., Benzer S., Rudnicka M., Miller C.A.;		
RL	"Calphotin: a Drosophila photoreceptor cell calcium-binding protein."		
CC	Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).		
CC	[2]		
RP	SEQUENCE FROM N.A.		
EX	STRAIN=Canton-S;		
RA	MEDLINE=93185730; PubMed=8434015;		
RT	Balling D.G., Xue N., Harshman K.D.;		
RL	"A Drosophila photoreceptor cell-specific protein, calphotin, binds		
CC	calcium and contains a leucine zipper."		
CC	Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).		
CC	-!- FUNCTION: Might function as a calcium-sequestering "sponge" to		
CC	regulate the amount of free cytoplasmic calcium. It binds 0.3 mole		
CC	of Ca(2+) per mole of protein.		
CC	-!- SUBUNIT: Homodimer (Probable).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic; hypodense compartment.		
CC	-!- TISSUE SPECIFICITY: Soma and axons of photoreceptor cells of		
CC	compound eyes and ocelli.		
CC	-!- DEVELOPMENTAL STAGE: Expressed early in photoreceptor cell		
CC	development.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		

15-MAR-2004 (Rel. 43, Last annotation update)
Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4.
HNCN4.
Rattus norvegicus (Rat).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
STRAIN=Sprague-Dawley; TISSUE=Brain;
MEDLINE=20457301; PubMed=11000485;
Monteggia L.M., Elsch A.J., Tang M.D., Kaczmarek L.K., Nestler E.J.;
"Cloning and localization of the hyperpolarization-activated cyclic
nucleotide-gated channel family in rat brain.";
Brain Res. Mol. Brain Res. 81:129-139(2000).
[2]
SEQUENCE OF 262-428 FROM N.A.
TISSUE=Heart;
MEDLINE=99459217; PubMed=10400919;
Shi W., Wymore R., Yu H., Wu J., Wymore R.T., Pan Z., Robinson R.B.,
Dixon J.E., McKinnon D., Cohen I.S.;
"Distribution and prevalence of hyperpolarization-activated cation
channel (HCN) mRNA expression in cardiac tissues.";
Circ. Res. 85:1-6(1999).
[3]
FUNCTION, AND TISSUE SPECIFICITY.
MEDLINE=21530492; PubMed=11675786;
Stevens D.R., Seifert R., Bufe B., Mueller F., Kremmer E., Gauss R.,
Meyerhof W., Kaupp U.B., Lindemann B.;
"Hyperpolarization-activated channels HCN1 and HCN4 mediate responses
to sour stimuli.";
Nature 413:631-635(2001).
CC CC -!- FUNCTION: Hyperpolarization-activated ion channel with very slow
activation and inactivation exhibiting weak selectivity for
potassium over sodium ions. May contribute to the native Pacemaker
currents in heart (If) and in neurons (Ih). Activated by CAMP. May
mediate responses to sour stimuli.
CC CC -!- SUBUNIT: The potassium channel is probably composed of a homo- or
heterotetrameric complex of pore-forming subunits.
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -!- TISSUE SPECIFICITY: Highly expressed in pyramidal and granule
layer of the hippocampus, in thalamus anterior nucleus, in the
supraoptic nucleus in hypothalamus, in cerebellum, and in
trapezoid nuclei and superior olivary complex in the auditory
system. Detected in a subset of elongated cells in taste buds.
CC CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
characterized by a series of positively charged amino acids at
every third position.
CC CC -!- SIMILARITY: Belongs to the potassium channel family. HCN
subclass.
CC CC -!- SIMILARITY: Contains 1 cyclic nucleotide-binding domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licenses@isb-sib.ch).

EMBL; AF247453; AAF62176.1; --
DR DR EMBL; AF155166; AAF01493.1; --
DR DR InterPro; IPR000595; CNMP_binding.
DR DR InterPro; IPR005821; Ion_trans.
DR DR InterPro; IPR005820; M+channel_nlg.
DR DR Pfam; PF00027; CNMP_binding; 1.
DR DR Pfam; PF00520; Ion_Trans; 1.
DR DR SMART; SM00100; CNMP; 1.
DR DR PROSITE; PS00888; CNMP_BINDING_1; 1.
DR DR PROSITE; PS00889; CNMP_BINDING_2; FALSE_NEG.
DR DR PROSITE; PS50042; CNMP_BINDING_3; 1.
KW Transport; Ion transport; Ionic channel; Voltage-gated channel;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:18:15 ; Search time 36.5 seconds

(without alignments)

2809.409 Million cell updates/sec

Title: US-10-720-192-3

Perfect score! 1726

Sequence: 1 MHQVDPNLTTRKGRALAI.....PAGEVAPTPTPTPQRTLPA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp Vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1539	89.2	286	Q9r640	mycobacteri
2	1053.5	61.0	368	Q9r4H9	mycobacteri
3	568	32.9	194	Q8V565	mycobacteri
4	256	14.8	3247	Q65553	bovine herp
5	248.5	14.4	698	Q9ASK4	criza sativ
6	248	14.4	3288	Q7T5D9	simian herp
7	244	14.1	511	Q95JD0	sus scrofa
8	244	14.1	566	Q95JD1	sus scrofa
9	241.5	14.0	745	Q89X06	bradyrhizob
10	240.5	13.9	3326	Q7T591	simian herp
11	238	13.8	1269	Q8W5K6	criza sativ
12	238	13.8	1269	Q7XH56	criza sativ
13	237.5	13.8	1188	Q41805	zea mays (m
14	231	13.4	668	Q8XU08	caenorhabdi
15	230.5	13.4	584	Q82D07	streptomyce
16	228.5	13.2	464	Q41645	volvox cart

17	227.5	13.2	552	4	Q8NAA1	Q8naal homo sapien
18	227	13.2	598	16	Q8VKN7	mycobacteri
19	226.5	13.1	222	10	Q9S740	arabidopsis
20	226.5	13.1	609	16	Q8FM53	corynebacte
21	225.5	13.1	420	5	Q9VZC2	drosophila
22	225.5	13.1	652	10	Q94JZ6	arabidopsis
23	225	13.0	676	6	Q95JC9	sus scrofa
24	224.5	13.0	3534	12	Q39266	equine herp
25	223.5	12.9	652	10	Q91V48	arabidopsis
26	222	12.9	763	2	Q9XDH2	mycobacteri
27	222	12.9	3084	12	Q8UZ11	pseudorabie
28	218	12.6	409	10	Q8SBM1	volvox cart
29	217.5	12.6	839	16	Q9RX57	deinococcus
30	217	12.6	784	13	Q90YB5	gallus gall
31	217	12.6	801	5	Q23635	caenorhabdi
32	216.5	12.5	496	16	Q7U8L6	synecococc
33	216.5	12.5	847	10	Q9XIB6	arabidopsis
34	215.5	12.5	817	3	Q07229	saccharomyc
35	214.5	12.4	667	10	Q948Y7	volvox cart
36	214.5	12.4	1016	10	Q9SPM0	zea mays (m
37	214	12.4	1204	3	Q8JOE5	yarrowia li
38	214	12.4	1627	10	Q84ZL0	oryza sativ
39	213.5	12.4	760	10	Q9TOK5	arabidopsis
40	213.5	12.4	3122	12	P89459	herpes simp
41	212	12.3	753	5	Q93107	acanthamoeb
42	211	12.2	889	16	Q9F2N5	streptomyce
43	209	12.1	1480	4	Q96Q04	homo sapien
44	208.5	12.1	214	10	Q93763	gossypium b
45	206.5	12.0	575	16	Q98F98	rhizobium l

ALIGNMENTS

RESULT 1

Q9r640 PRELIMINARY; PRT; 286 AA.
 AC Q9r640;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE MPT-32=45 kDa CULTURE FILTRATE glycoprotein.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_taxid=1773;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=96196153; PubMed=86263147;
 RA Dobos K.M., Khoo K.H., SwiderEK K.M., Brennan P.J., Belisle J.T.;
 RT "Definition of the full extent of glycosylation of the 45-kilodalton
 glycoprotein of Mycobacterium tuberculosis."
 RL J. Bacteriol. 178:2498-2506(1996).
 SQ SEQUENCE 286 AA; 28764 MW; 4BD7AD89FEB42493 CRC64;

Query Match	89.2%	Score 1539;	DB 2;	Length 286;
Best Local Similarity	99.7%	Pred No. 1.3e-89;		
Matches 285;	Conservative	17 Mismatches	0; Indels	0; Gaps
Cy	40	DPBPAPVPTTAASPSSTAAAPAPATPVAPPPAAANTPNAOPGDPNAAAPPADNAPP	99	
Db	1	DPBPAPVPTTAASPSSTAAAPAPATPVAPPPAAANTPNAOPGDPNAAAPPADNAPP	60	
Qy	100	PPVIAPNAPQVRIDNPVGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGPPPV	159	
Db	61	PPVIAPNAPQVRIDNPVGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGPPPV	120	
Qy	160	ANDTRIVLGRDOKLYASAEATDSKAAALGSDMGFFYMPYPTGTRINQETVSLDANGVG	219	
Db	121	ANDTRIVLGRDOKLYASAEATDSKAAALGSDMGFFYMPYPTGTRINQETVSLDANGVG	180	
Qy	220	SASVVEKESDKPKNGQIWTGVIQSPANAPDAGPPQRFVVLGTANNPVDKGAAL	279	

RP	SEQUENCE FROM N.A.
RC	STRAIN=cv. Nipponbare;
RA	Sasaki T., Matsumoto T., Yamamoto K.;
RN	"Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT	clone:P0439B06.";
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC	-/- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR	EMBL; AP002882; BAB39873.1; ..
DR	Gramene; Q9ASK4; ..
DR	GO: GO:0005524; F:ATP binding; IEA.
DR	GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR	GO: GO:0016740; F:transferase activity; IEA.
DR	GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR	InterPro; IPR000719; Prot_kinase.
DR	InterPro; IPR008271; Ser_thr_pkin_AS.
DR	Pfam; PF00069; pkinase_1.
DR	ProDom; PD00001; Prot_kinase; 1.
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR	PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR	PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR	KW Hypothetical protein; ATP-binding; Kinase;
KW	Hypothetical protein; ATP-binding; Kinase;
KW	Serine/threonine-protein kinase; Transferase.
QZ	SEQUENCE 698 AA; 72733 MW; F3866E46DCA9DA56 CRC64;
Query Match	14.4%; Score 248.5; DB 10; Length 698;
Best Local Similarity	30.7%; Pred.No. 7.5e-08;
Matches 89; Conservative	20; Mismatches 134; Indels 47; Gaps 9;
QY	32 AVPATANADPEAPPPVETTAASDPSTAAAPPAPATPVAPPPPAAPAATPNACQGDPAAPP 91
Db	12 ATATAAASPSTATPTP--ATGNATPADPSITPPAASPLPSAAITPPQDSPSPSP 68
QY	92 PADNPAPPPVLAPNAPQVRINDVPNGGFSFALPAGWVESDAHFYGSALLSKTTGGPP 151
Db	69 PAPPVAVPPPATVPVPPVPVASPPPT-----PSATLPSPSPASVPVPTPATTPPKP 121
QY	152 FPGQPFPVANDTRIVLGRLDKLYASAETAATSKAAAEGLGDMGEFYM-PYPGTIRINGTV 210
Db	122 SPVQQPFVA-----ASPPSSPADLPFPNPPAR--SDTPPVQSPPPPhRRSRPTP 170
QY	211 SLDANGVSGASYVEKFSKPSKPNGIWTGVIGSPAANADAGPPQRWFVWLGTANNP 270
Db	171 WAPMPAPSPTS-----PTKP-----SPASPSPIAGDP-----IIPTPNP 205
QY	271 VDGAKAALAESIRPLVAPPAPAPAEPAPAPAGEVAPTPTTTPQ 320
Db	207 SSPLATPSACSGTGTPVT-PSAPVSGDPSPTGATATAADRNSKLSENTQ 255
RESULT 6	
Q7T5D9	PRELIMINARY;
ID Q7T5D9	PRT; 3288 AA.
AC Q7T5D9	
DT 01-OCT-2003	(TrEMBLrel. 25, Created)
DT 01-OCT-2003	(TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003	(TrEMBLrel. 25, Last annotation update)

OS	Smian herpes B virus (Cercopithecid herpesvirus 1) (Shbv).
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC	Alphaherpesvirinae; Simplexvirus.
NCBI_TaxID=10325;	
[1]	
RP	SEQUENCE FROM N.A.
RC	STRAIN=E2490;
RX	MEDLINE=22628476; PubMed=12743273;
RA	Perelygina L., Zhu L., Zurkulan H., Mills R., Borodovsky M.,
RA	Hilliard J.K.;
RT	"Complete Sequence and Comparative Analysis of Herpes B
RT	Virus (Cercopithecine Herpesvirus 1) from a Rhesus Monkey.";
RL	J. Virol. 77:6167-6177(2003).
DR	EMBL; AF533768; AAP41454.1; -;
SO	SEQUENCE 3288 AA; 1342496 MW; 79C391EE856F7983 CRC64;

	Query Match	14.4%	Score 248;	DB 12; Length 3288;
	Best Local Similarity	29.8%;	Pred. No.	4.4e-07;
	Matches	91; Conservative	24; Mismatches	150; Indels 40; Gaps 9;
QY	30 TVAVPATANADPPAPDPVPTTAASPSTAAAPAPATPVAPPAAAAANTNAQGDFNAA	89		
Db	: : : :	:	:	:
	2853 TVAAGSGASAPPAPAFAAFAAAAAPAAPAAPAAPAAPAAPAAP-	AAPAPAAP	2511	
QY	90 PPAADNPAPPPV--IAPNAPQVRIDNPVGGSFALPAGWVSDDAHFDYGALLSKTT	147		
Db		:	:	:
	2912 AATAAFAAFAPAAPAAPAAPAAVAVVVPAPTATITATTTSAPAAATPASFPVTPTS	2971		
QY	148 GDPPFGQP-----PPVANDTRIVLG----RLDQLKYSAEATDSKAAR----	LGS 191		
Db	: : : :	:	:	:
	2972 SLTPSPKPFATOPSLATGCGSVAPGDFFRRAPSRTAAVPAAPSRPPARRLARPAVS	R 3031		
QY	192 DMGEFYMPYPETGRINOTVSLDANGVSGSASYEYVKFSDFPSKPNGOIQTWGVISGAANAP	251		
Db	: : : :	:	:	:
	3032 STSFALP-PDELARPRTPEAPPTETEENAPVAERPAPPEPGQR-----PPSPAAP	3083		
QY	252 DAGGPQEWFVMGLTANNPYDKGNAXALAESIRPLVAPP-----PAPAPAEAEPAPAPA	307		
Db	: : : :	:	:	:
	3084 DAGE-----AAASGPSGGVPAPRELVALPGRVAVPRRQQIPPPAPPREIFAPSPPPP	3134		
QY	308 GEVAP 312			
Db				
	3135 RSHAP 3139			
RESULT 7				
Q95JDO	PRELIMINARY;	PRT;	511 AA.	
ID	Q95JDO PRELIMINARY;			
AC	Q95JDO;			
DT	01-DRC-2001 (TrEMBLrel_19, Created)			
DT	01-DRC-2001 (TrEMBLrel_19, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel_25, Last annotation update)			
DE	Basic proline-rich protein.			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;			
NCBI_TaxID=9823;	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
SEQUENCE FROM N.A.				
TISSUE=Pituitary gland;				
Zhang Q., Szalay A.A., Kyeuyne-Nyombi E., Sands J.F., Oberg K.C.,				
Tieche J.-M., Leonora J.;				
"Cloning and expression of a novel proline-rich protein from porcine parotid glands,"				
Submitted (May-2001) to the EMBL/GenBank/DDBJ databases.				
EWEL; AY035648; AK61382.1; -				
G; GO:0005199; Fstructural constituent of cell wall; IEA.				
IinterPro; IPR003882; Pistil extensin.				
D PRINTS; PR01218; PSTLEXTENSIN.				
DR SEQUENCE 511 AA; 48483 MW; AB04597964C448D7 CRC64;				

Query Match	14.1%	Score 244;	DB 6;	Length 511;
Best Local Similarity	27.9%;	Pred. No. 1e-07;		
Matches	81;	Conservative	10;	Mismatches 99; Indels 100; Gaps 13;
QY	34	PATANADPERAPPVPTTAASPSSTAAPPAPATPVAPPAAAAANTNAOGDENAAPP	93	
Db	72	PAPPGARPPPGPPP----	GPDPGAPPGARPPGPPPP----	GPDPGAPPGARPPPG 124
QY	94	-DENAPPPEVIAPNAPQVRINDNVGGSFALPAGWVSDAAHFYGSALLSKTTGDDPP-	151	
Db	125	PPPPGPPPGAPPGARPPGPPPP----	PAGGLQQGPAPSHVG----	PKKKPPPPG 172
QY	152	--FFCQPPPVANDTRIVLGRLDQKLYASAEATDSKAAARLGSMDGSEFYMPYCTRINQET	209	
Db	173	AGHPRPPPPPPAN-----	ESQGPGR-----	PPPG----- 195
QY	210	VSLDANGVSGSAYVEVKFSDSPKPNQGOIWTGIGSPAAANAPDAGPPORFWFVWLGTANN	269	

```
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iiguchi M., Kawashima K.,
RA Kohara Y., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005936; BAC45786.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004672; F:protein kinase activity; IEA.
DR GO; GO:0005199; F:structural constituent of cell wall; IEA.
DR GO; GO:0006488; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR006665; OmpA; IEA.
DR InterPro; IPR003882; Pistil extensin.
DR InterPro; IPR000719; Prot_Kinase.
DR Pfam; PF00691; OmpA; 1.
DR PRINTS; PR01218; PSTLEXKINSE.
DR ProDom; PD000001; Prot_Kinase; 1.
DR Complete proteome.
KW SEQUENCE 745 AA; 74545 MW; 155EDFC74BDF6D CRC64;

Query Match 14.0%; Score 241.5; DB 16; Length 745;
Best Local Similarity 29.6%; Pred. No. 2.2e-07;
Matches 112; Conservative 16; Mismatches 136; Indels 115; Gaps 21;

QY 15 LAALAIAN-----ASASLVTVAVPATANADPEP-----APP----- 46
DB 9 LATTALTANQLANTASHAQGASPLVVAQAQOETGPDGKQPKPPGGAAPPAARPA 68
QY 47 VPTTAASPSSTAAPP---APATVAPPPPAANTNAQGDPAAPDPNAP-PPPV 102
DB 69 APPAAAPHPHPPAPPAAPRAAPPPPP-PPAARAPPPPPPPPPAACKQSPPPA 126
QY 103 IAP--NAPQVRIIDNVGSGFSAFPAGWVESDAHFYGSALLSKTTGDP-PFPGQPPV 159
DB 127 AAPQQAFTP---PPAPPAARPAFTPPAPPAAPQAHAPPPPPPPAARPTPTPPPPPA 183
QY 160 ANDRIVLGRDQKLYASAEATDSKAA-----ARLGSMDGEFYMYFG-----TRI 205
DB 184 GPAARPT-----PAPTATPTFVAPPAAPTARPGS-----PAPAATPAFTTTPA 227
QY 206 NOETVSLDANGVSGSASYEYKFSKPDSPKPNQIWTGVIGSPAANAPAGPPQRFVVMWL 265
DB 228 PTATPAFTATAPGSTP-----GAPPAGPGAPPG-----VRPG 262
QY 266 T---ANNPVDKGAALAESTRP--LVAPP-----PA--PAPAPAEAP--APAPAGEVA 311
DB 263 SPPAAGSPAPGATPAFTTTTAPGATATPSPGPGPASTPAPGATPAFTATPAPGGALT 322
QY 312 PTP-----TPTPTORTLP 324
DB 323 PPGPGAGPTPGPGGTP 341

RESULT 10
QY7591 PRELIMINARY; PRT; 3326 AA.
AC QY7591;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Large tegument protein.
GN UL36.
OS Simian herpes B virus (Cercopithecoid herpesvirus 1) (Shbv).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10325;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22607624; PubMed=12721804;

RA Ohsawa K., Black D.H., Sato H., Rogers K., Eberle R.;
RT "Sequence and genetic arrangement of the UL region of the monkey B
RT virus (Cercopithecine herpesvirus 1) genome and comparison with the UL
RT region of other primate herpesviruses.";
RL Arch. Virol. 148:989-997(2003).
DR EMBL; AB096160; BAC58076.2; -.
SQ SEQUENCE 3326 AA; 345566 MW; 6B53E3860F43CDF0 CRC64;

Query Match 13.9%; Score 240.5; DB 12; Length 3326;
Best Local Similarity 29.1%; Pred. No. 1.3e-06;
Matches 93; Conservative 24; Mismatches 162; Indels 41; Gaps 9;

QY 17 ALAIAWASASLVTVAVPATANADPEPAPVPPTTAASPPSTAAAPAPATPVAP--PPPA 74
DB 2875 AAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAPAA 2934
QY 75 AANTPAQFQDNAPPPADPNAPPPV--IAPNAPQVRIIDNVGSGFSAFPAGWVESD 132
DB 2935 AVPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAAPAA 2994
QY 133 AAHFDYGSALLSKTTGDPFPGQP-----PPVANDTRIVLG-----RLDQKLYASAEATD 182
DB 2995 APAATPASVPPTTSSLPTTSPKPPAFFQFSLATGSGVAPGDFRRRAPSRPTAAVPAAP 3054
QY 183 SKAAAR-----LGSDMGFEFYMPYPGTRINQETVSLDANGVSGSASYEYKFSKPDSPKPN 236
DB 3055 SRPPARLARPAVSRSTESFALP-PDELARPTPEAPAPTETEAPVAERPAPEPPQG 3113
QY 237 QIWTGVIGSPAANAPAGPPQRFVVMWLGTANNPVDKGAALAEISIRLVAPP-----PA 292
DB 3114 R-----PPSPAAPDAGP-----AAASGSGVGPAPRLGALVPGRVAPRRQIPP 3157
QY 293 PAPAPAEAPAPAPAGEVAP 312
DB 3158 PAPPREIPAPPPPPRSHAP 3177

RESULT 11
QY8W5K6 PRELIMINARY; PRT; 1269 AA.
AC QY8W5K6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OSJNBA0079B05.10 OR OSJNAA0079B05.2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Sasaki C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC079179; AAL31655.1; -.
DR EMBL; AC116601; AAM08709.1; -.
DR Gramene; Q8W5K6; -.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR008973; C2.CaLB.
DR InterPro; IPR003104; FH2.
DR Pfam; PF02181; FH2; 1.
DR SMART; SM00498; FH2; 1.
DR Hypothetical protein.
KW SEQUENCE 1269 AA; 138432 MW; 262E546481B25CA6 CRC64;
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:21:10 ; Search time 16.5 Seconds
(without alignments)
1016.875 Million cell updates/sec

Title: US-10-720-192-3
Perfect score: 1726
Sequence: 1 MHQVDNLTTRKGRLAALAI.....PAGEVAPTTPTPQRTLPA 325

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/FACTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1726	100.0	325	1	US-08-382-184-2
2	1726	100.0	325	2	US-08-641-356-2
3	1726	100.0	325	3	US-09-132-528-2
4	1726	100.0	325	3	US-09-132-528-3
5	1726	100.0	325	3	US-08-875-494-2
6	1726	100.0	325	4	US-09-599-366-2
7	1726	100.0	325	4	US-09-599-366-3
8	1726	100.0	325	4	US-08-875-494-2
9	1726	100.0	325	4	US-09-510-031A-6
10	1726	100.0	325	4	US-09-985-372-2
11	1726	100.0	325	4	US-09-985-372-3
12	1726	100.0	332	3	US-08-818-111-53
13	1726	100.0	332	4	US-08-818-111-53
14	1726	100.0	332	4	US-09-056-556-53
15	1726	100.0	332	4	US-09-072-596-53
16	1726	100.0	332	4	US-09-072-967-53
17	1720	99.7	325	4	US-09-510-031A-2
18	1542	89.3	286	1	US-08-382-184-3
19	1542	89.3	286	2	US-08-641-356-3
20	1542	89.3	286	3	US-09-132-528-4
21	1542	89.3	286	3	US-08-875-494-3
22	1542	89.3	286	4	US-09-599-366-4
23	1542	89.3	286	4	US-08-875-494-3
24	1542	89.3	286	4	US-09-985-372-4
25	1539	89.2	286	4	US-09-510-031A-7
26	1499	86.8	652	4	US-09-072-596-350
27	1499	86.8	652	4	US-09-072-967-355

28	1499	86.8	802	4	US-09-056-556-214	Sequence 214, App
29	1499	86.8	802	4	US-09-072-596-209	Sequence 209, App
30	1499	86.8	802	4	US-09-072-596-346	Sequence 346, App
31	1499	86.8	802	4	US-09-072-967-214	Sequence 214, App
32	1499	86.8	802	4	US-09-072-967-351	Sequence 351, App
33	1499	86.8	802	4	US-09-287-849-10	Sequence 10, Appl
34	985	57.1	381	4	US-09-510-031A-5	Sequence 5, Appl
35	980	56.8	287	4	US-08-311-731A-96	Sequence 96, Appl
36	980	56.8	287	4	US-09-510-031A-4	Sequence 4, Appl
37	969	56.1	287	4	US-09-510-031A-8	Sequence 8, Appl
38	233	13.5	8991	4	US-08-714-741-32	Sequence 32, Appl
39	200.5	11.6	214	1	US-08-217-327-4	Sequence 4, Appl
40	200	11.6	408	1	US-07-609-716-65	Sequence 65, Appl
41	200	11.6	408	3	US-08-475-411A-65	Sequence 65, Appl
42	200	11.6	408	3	US-08-478-029A-65	Sequence 65, Appl
43	197.5	11.4	330	1	US-08-642-255-32	Sequence 32, Appl
44	197	11.4	538	4	US-09-616-289-43	Sequence 43, Appl
45	196.5	11.4	666	4	US-09-050-739-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1

US-08-382-184-2
; Sequence 2, Application US/08382184
; Patent No. 5714593

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: MICROBACTERIAL PROTEINS,

; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINES
; AND FOR THE DETECTION OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &

; ADDRESS: NEUSTADT, P.C.

; STREET: 1755 S. Jefferson Davis Highway, Suite 400

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/382,184

; FILING DATE: 01-FEB-1995

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: OBLON, NORMAN F

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 22640720

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-413-3000

; TELEFAX: 703-413-2220

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 325 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-382-184-2

Query Match 100.0%; Score 1726; DB 1; Length 325;

Best Local Similarity 100.0%; Pred No. 6.4e-127;

Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 MHQVDNLTTRKGRLAALAIAMASASLVTVAVPATANADPEFPVPTTAASPFSTAAA 60

1 MHQVDNLTTRKGRLAALAIAMASASLVTVAVPATANADPEFPVPTTAASPFSTAAA 60

Db

QY 61 PPAPATVPAPPPAAANTPNAQGDPAAPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
 Db 61 PPAPATVPAPPPAAANTPNAQGDPAAPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
 QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDDPPPGPPGVANDTRIVLGRDOKLYASAEA 180
 Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDDPPPGPPGVANDTRIVLGRDOKLYASAEA 180
 QY 181 TDSKAAARLGSDMGFEFYPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
 Db 181 TDSKAAARLGSDMGFEFYPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
 QY 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEISIRPLVAPPPAPAPAPAE 300
 Db 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEISIRPLVAPPPAPAPAPAE 300
 QY 301 APAPAPAGEVAPTPTTPTQRTPLA 325
 Db 301 APAPAPAGEVAPTPTTPTQRTPLA 325

RESULT 2
 US-08-641-356-2
 ; Sequence 2, Application US/08641356
 ; Patent No. 5866130
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: MICROBACTERIAL PROTEINS,
 ; TITLE OF INVENTION: MICROORGANISMS PRODUCING THEM AND THEIR USE FOR VACCINES
 ; TITLE OF INVENTION: AND FOR THE DETECTION OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &
 ; ADDRESS: NEUSTADT, P.C.
 ; STREET: 1755 S. Jefferson Davis Highway, Suite 400
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/641,356
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/382,184
 ; FILING DATE: 01-FEB-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OBLON, NORMAN F
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 22640720
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-413-3000
 ; TELEFAX: 703-413-2220
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 325 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-641-356-2

Query Match 100.0%; Score 1726; DB 2; Length 325;
 Best Local Similarity 100.0%; Pred. No. 6.4e-127;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHQVDNLTFRKGRLLAALAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
 Db 1 MHQVDNLTFRKGRLLAALAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60

QY 61 PPAPATVPAPPPAAANTPNAQGDPAAPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
 Db 61 PPAPATVPAPPPAAANTPNAQGDPAAPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
 QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDDPPPGPPGVANDTRIVLGRDOKLYASAEA 180
 Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDDPPPGPPGVANDTRIVLGRDOKLYASAEA 180
 QY 181 TDSKAAARLGSDMGFEFYPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
 Db 181 TDSKAAARLGSDMGFEFYPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
 QY 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEISIRPLVAPPPAPAPAPAE 300
 Db 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEISIRPLVAPPPAPAPAPAE 300
 QY 301 APAPAPAGEVAPTPTTPTQRTPLA 325
 Db 301 APAPAPAGEVAPTPTTPTQRTPLA 325

RESULT 3
 US-09-132-528-2
 ; Sequence 2, Application US/09132528A
 ; Patent No. 6221353
 ; GENERAL INFORMATION:
 ; APPLICANT: LAQUEYREIRB, Anne
 ; APPLICANT: MARCHAL, Gilles
 ; APPLICANT: PESCHER, Pascale
 ; APPLICANT: ROWAIN, Felix
 ; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICORORGANISMS PRODUCING THEM
 ; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
 ; TITLE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 0660-0138-0DIV
 ; CURRENT APPLICATION NUMBER: US/09/132,528A
 ; CURRENT FILING DATE: 1998-08-11
 ; EARLIER APPLICATION NUMBER: 08/641,356
 ; EARLIER FILING DATE: 1996-04-30
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 325
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-132-528-2

Query Match 100.0%; Score 1726; DB 3; Length 325;
 Best Local Similarity 100.0%; Pred. No. 6.4e-127;
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHQVDNLTFRKGRLLAALAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
 Db 1 MHQVDNLTFRKGRLLAALAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
 QY 61 PPAPATVPAPPPAAANTPNAQGDPAAPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
 Db 61 PPAPATVPAPPPAAANTPNAQGDPAAPADPNAPPPVIAFNAPQPVRIIDNPVGGF 120
 QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDDPPPGPPGVANDTRIVLGRDOKLYASAEA 180
 Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDDPPPGPPGVANDTRIVLGRDOKLYASAEA 180
 QY 181 TDSKAAARLGSDMGFEFYPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
 Db 181 TDSKAAARLGSDMGFEFYPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
 QY 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEISIRPLVAPPPAPAPAPAE 300
 Db 241 GVIGSPAANAPDAGPPQRFVWVLTANNPVDKGAALAEISIRPLVAPPPAPAPAPAE 300
 QY 301 APAPAPAGEVAPTPTTPTQRTPLA 325
 Db 301 APAPAPAGEVAPTPTTPTQRTPLA 325

```

; EARLIER FILING DATE: 1996-02-01
; EARLIER APPLICATION NUMBER: 382184
; EARLIER FILING DATE: 1995-02-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-08-875-494-2

Query Match      100.0%; Score 1726; DB 3; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDNLTRRKGRLLAALAIAAMASASLTVAVPATANADPEPAPPVPTTAAASPSTAAA 60
Db 1 MHQVDNLTRRKGRLLAALAIAAMASASLTVAVPATANADPEPAPPVPTTAAASPSTAAA 60

Qy 61 PPAPATVPAPPPPPAAANTNNAOQDNPAPPPADPNAPPPPVTAAPNAQPVRIIDNPVGGF 120
Db 61 PPAPATVPAPPPPPAAANTNNAOQDNPAPPPADPNAPPPPVTAAPNAQPVRIIDNPVGGF 120

Qy 121 SFALPAGWVESDAAHFDYGSALLSKITGTGDPFFQGGPPPVANDTRIVLGRLDQKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKITGTGDPFFQGGPPPVANDTRIVLGRLDQKLYASAEA 180

Qy 181 TDSKAARLGSDMGCEYMPYPGTRINQETVSLDANGVSGSASYEYKFSDPSPNGQIWT 240
Db 181 TDSKAARLGSDMGCEYMPYPGTRINQETVSLDANGVSGSASYEYKFSDPSPNGQIWT 240

Qy 241 GVIGSPAANAPDAGPQRFVVMWLTANPNVDKGAAKALAEISRPLVAPPAPAPAPAPAE 300
Db 241 GVIGSPAANAPDAGPQRFVVMWLTANPNVDKGAAKALAEISRPLVAPPAPAPAPAPAE 300

Qy 301 APAPAPAGEVAPPTTPTPTQRTPLA 325
Db 301 APAPAPAGEVAPPTTPTPTQRTPLA 325

RESULT 6
US-09-599-366-2
; Sequence 2, Application US/09599366
; Patent No. 6335181
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: VARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/599,366
; CURRENT FILING DATE: 2000-06-21
; PRIOR APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: 1998-08-11
; PRIOR APPLICATION NUMBER: 08/641,356
; PRIOR FILING DATE: 1996-04-30
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-599-366-2

Query Match      100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDNLTRRKGRLLAALAIAAMASASLTVAVPATANADPEPAPPVPTTAAASPSTAAA 60

```

Db 1 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
 Qy 61 PPAPATPVAPPVPPAAANTENACPDNAAAPPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
 Db 61 PPAPATPVAPPVPPAAANTENACPDNAAAPPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
 Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGPPVAVNDTRIVLGRDOKLYASAEA 180
 Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGPPVAVNDTRIVLGRDOKLYASAEA 180
 Qy 181 TDSKAARLGSDMGGEFYPGTRINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240
 Db 181 TDSKAARLGSDMGGEFYPGTRINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240
 Qy 241 GVIGSPAANAPDAGPPQRFVVMGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAP 300
 Db 241 GVIGSPAANAPDAGPPQRFVVMGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAP 300
 Qy 301 APAPAGEVAPTPTTPTPQRTLPA 325
 Db 301 APAPAGEVAPTPTTPTPQRTLPA 325

RESULT 7

US-09-599-366-3

; Sequence 3, Application US/09599366

; Patent No. 6335181

; GENERAL INFORMATION:

; APPLICANT: LAQUEYRIERIE, Anne

; APPLICANT: MARCHAL, Gilles

; APPLICANT: PESCHER, Pascale

; APPLICANT: ROMAIN, Felix

; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM

; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF

; FILE REFERENCE: 0660-0138-ODIV

; CURRENT APPLICATION NUMBER: US/09/599,366

; PRIOR FILING DATE: 2000-06-21

; PRIOR FILING DATE: 1998-08-11

; PRIOR FILING DATE: 1998-04-30

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 325

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-09-599-366-3

Query Match 100.0%; Score 1726; DB 4; Length 325;

Best Local Similarity 100.0%; Pred. No. 6.4e-127;

Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60

Db 1 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60

Qy 61 PPAPATPVAPPVPPAAANTENACPDNAAAPPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120

Db 61 PPAPATPVAPPVPPAAANTENACPDNAAAPPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120

Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGPPVAVNDTRIVLGRDOKLYASAEA 180

Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGPPVAVNDTRIVLGRDOKLYASAEA 180

Qy 181 TDSKAARLGSDMGGEFYPGTRINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240

Db 181 TDSKAARLGSDMGGEFYPGTRINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240

Qy 241 GVIGSPAANAPDAGPPQRFVVMGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAP 300

Db 241 GVIGSPAANAPDAGPPQRFVVMGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAP 300

Qy 301 APAPAGEVAPTPTTPTPQRTLPA 325
 Db 301 APAPAGEVAPTPTTPTPQRTLPA 325

RESULT 8

US-08-875-494-2

; Sequence 2, Application US/08875494

; Patent No. 6379902

; GENERAL INFORMATION:

; APPLICANT: LAQUEYRIERIE, ANNE

; APPLICANT: MARCHAL, GILLES

; APPLICANT: PESCHER, PASCALE

; APPLICANT: ROMAIN, FELIX

; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM

; TITLE OF INVENTION: AND THEIR USES FOR VACCINES AND FOR THE DETECTION OF

; FILE REFERENCE: 0660-0122-0 PCT

; CURRENT APPLICATION NUMBER: US/08/875,494

; CURRENT FILING DATE: 1997-08-01

; PRIOR APPLICATION NUMBER: PCT/FR96/00166

; PRIOR FILING DATE: 1996-02-01

; PRIOR APPLICATION NUMBER: 382184

; PRIOR FILING DATE: 1995-02-01

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2

; LENGTH: 325

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; US-08-875-494-2

Query Match 100.0%; Score 1726; DB 4; Length 325;

Best Local Similarity 100.0%; Pred. No. 6.4e-127;

Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60

Db 1 MHQVDPNLTTRKGRLLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60

Qy 61 PPAPATPVAPPVPPAAANTENACPDNAAAPPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120

Db 61 PPAPATPVAPPVPPAAANTENACPDNAAAPPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120

Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGPPVAVNDTRIVLGRDOKLYASAEA 180

Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPDPGPPVAVNDTRIVLGRDOKLYASAEA 180

Qy 181 TDSKAARLGSDMGGEFYPGTRINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240

Db 181 TDSKAARLGSDMGGEFYPGTRINQETVSLDANGVSGSASYEVKFSDFSKNGQIWT 240

Qy 241 GVIGSPAANAPDAGPPQRFVVMGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAP 300

Db 241 GVIGSPAANAPDAGPPQRFVVMGTANNPVDKGAAKALAESIRPLVAPPAPAPAPAP 300

Qy 301 APAPAGEVAPTPTTPTPQRTLPA 325

Db 301 APAPAGEVAPTPTTPTPQRTLPA 325

RESULT 9

US-09-510-031A-6

; Sequence 6, Application US/09510031A

; Patent No. 6638518

; GENERAL INFORMATION:

; APPLICANT: Ratliff, Timothy

; APPLICANT: Kline, Joel

; TITLE OF INVENTION: METHOD FOR INHIBITING INFLAMMATORY RESPONSES

; FILE REFERENCE: 140.0010.0101

; CURRENT APPLICATION NUMBER: US/09/510,031A

; CURRENT FILING DATE: 2000-02-22

```
; PRIOR APPLICATION NUMBER: US 60/121,177
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 6
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-510-031A-6

Query Match      100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
Db 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
Qy 61 PPAPATPVAPPPPPAAAANTPNAQPGDPNAAAPPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPPPPAAAANTPNAQPGDPNAAAPPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDOKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDOKLYASAEA 180
Qy 181 TDSKAAARLGSDMGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
Db 181 TDSKAAARLGSDMGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
Qy 241 GVIGSPAANADAGPPQRFVWVLTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 300
Db 241 GVIGSPAANADAGPPQRFVWVLTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 300
Qy 301 APAPAPAGEVAPTPTTPTPQRTLEA 325
Db 301 APAPAPAGEVAPTPTTPTPQRTLEA 325

RESULT 10
US-09-985-372-2
; Sequence 2, Application US/09985372
; Patent No. 6676945
; GENERAL INFORMATION:
; APPLICANT: LAQUEYRERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 3
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-3

Query Match      100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
Db 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
Qy 61 PPAPATPVAPPPPPAAAANTPNAQPGDPNAAAPPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
Db 61 PPAPATPVAPPPPPAAAANTPNAQPGDPNAAAPPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDOKLYASAEA 180
Db 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPVANDTRIVLGRDOKLYASAEA 180
Qy 181 TDSKAAARLGSDMGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
Db 181 TDSKAAARLGSDMGFEFMPYPGTRINQETVSLDANGVSGSASYEYVKFSDPSKPNQIWT 240
Qy 241 GVIGSPAANADAGPPQRFVWVLTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 300
Db 241 GVIGSPAANADAGPPQRFVWVLTANNPVDKGAALKALAESIRPLVAPPAPAPAPAP 300
Qy 301 APAPAPAGEVAPTPTTPTPQRTLEA 325
Db 301 APAPAPAGEVAPTPTTPTPQRTLEA 325

RESULT 10
US-09-985-372-2
; Sequence 2, Application US/09985372
; Patent No. 6676945
; GENERAL INFORMATION:
; APPLICANT: LAQUEYRERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-2

Query Match      100.0%; Score 1726; DB 4; Length 325;
Best Local Similarity 100.0%; Pred. No. 6.4e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
Db 1 MHQVDPNLTTRKGLAALAIAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
Qy 61 PPAPATPVAPPPPPAAAANTPNAQPGDPNAAAPPADPNAPPPPPVIAPNAPQPVRIIDNPVGGF 120
```

RESULT 12
US-08-818-112-53
; Sequence 53, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 133
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-53

Query Match 100.0%; Score 1726; DB 3; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDPNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
DB 8 MHQVDPNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
QY 61 PPAPATPVAPPPAAANTNAPGPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
DB 68 PPAPATPVAPPPAAANTNAPGPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGPPGVANDTRIVLGRDQKLYASAEA 180
DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGPPGVANDTRIVLGRDQKLYASAEA 187
QY 181 TDSKAARLGSDMGFYMYPGTRINQETVSLDANGVSGSASYEVKFSKPNQIWT 240
DB 188 TDSKAARLGSDMGFYMYPGTRINQETVSLDANGVSGSASYEVKFSKPNQIWT 247
QY 241 GVIGSPAANADAGPPQRFVWVWLGTANNPVDKGAALAESIRPLVAPPPAPAPAPAE 300
DB 248 GVIGSPAANADAGPPQRFVWVWLGTANNPVDKGAALAESIRPLVAPPPAPAPAPAE 307
QY 301 APAPAPAGEVAPPTPTPTQRTPLA 325
|||||

Db 308 APAPAPAGEVAPPTPTPTQRTPLA 332
RESULT 13
US-08-818-111-53
; Sequence 53, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-53

Query Match 100.0%; Score 1726; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHQVDPNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
DB 8 MHQVDPNLTTRKGRLAALAIAMASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
QY 61 PPAPATPVAPPPAAANTNAPGPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 120
DB 68 PPAPATPVAPPPAAANTNAPGPNAPADPNAPPPVPIAPNAPQPVRIIDNPVGGF 127
QY 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGPPGVANDTRIVLGRDQKLYASAEA 180
DB 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPGPPGVANDTRIVLGRDQKLYASAEA 187
QY 181 TDSKAARLGSDMGFYMYPGTRINQETVSLDANGVSGSASYEVKFSKPNQIWT 240
DB 188 TDSKAARLGSDMGFYMYPGTRINQETVSLDANGVSGSASYEVKFSKPNQIWT 247
QY 241 GVIGSPAANADAGPPQRFVWVWLGTANNPVDKGAALAESIRPLVAPPPAPAPAPAE 300
DB 248 GVIGSPAANADAGPPQRFVWVWLGTANNPVDKGAALAESIRPLVAPPPAPAPAPAE 307
QY 301 APAPAPAGEVAPPTPTPTQRTPLA 325
|||||

Db 308 APAPAPAGEVAPTPTTPTPQRTLPA 332
|||||
RESULT 14
US-09-056-556-53
; Sequence 53, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 622-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-53
Query Match 100.0%; Score 1726; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHQVDPNLTTRKGRLLAALAAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
Db 8 MHQVDPNLTTRKGRLLAALAAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
Qy 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAPPPADPNAPPPVPIAPNAPQPVRIENPVGGF 120
Db 68 PPAPATPVAPPPPPAAANTPNAQPGDPNAPPPADPNAPPPVPIAPNAPQPVRIENPVGGF 127
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRLDQKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRLDQKLYASAEA 187
Qy 181 TDSKAAARLGSDMGGEFYMPYGTTRINQETVSLDANGVSGSASYEVKFSKPKNGQIWT 240
Db 188 TDSKAAARLGSDMGGEFYMPYGTTRINQETVSLDANGVSGSASYEVKFSKPKNGQIWT 247
Qy 241 GVIGSPAANADAGPPQRFVWVLGTANNPVVDKGAALAEISIRPLVAPPAPAPAPAP 300
Db 248 GVIGSPAANADAGPPQRFVWVLGTANNPVVDKGAALAEISIRPLVAPPAPAPAPAP 307
Qy 301 APAPAPAGEVAPTPTTPTPQRTLPA 325
Db 308 APAPAPAGEVAPTPTTPTPQRTLPA 332

RESULT 15
US-09-072-596-53
; Sequence 53, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-53

Query Match 100.0%; Score 1726; DB 4; Length 332;
Best Local Similarity 100.0%; Pred. No. 6.5e-127;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MHQVDPNLTTRKGRLLAALAAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 60
Db 8 MHQVDPNLTTRKGRLLAALAAIAAASASLVTVAVPATANADPEPAPVPTTAASPPSTAAA 67
Qy 61 PPAPATPVAPPPPPAAANTPNAQPGDPNAPPPADPNAPPPVPIAPNAPQPVRIENPVGGF 120
Db 68 PPAPATPVAPPPPPAAANTPNAQPGDPNAPPPADPNAPPPVPIAPNAPQPVRIENPVGGF 127
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRLDQKLYASAEA 180
Db 128 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPPVANDTRIVLGRLDQKLYASAEA 187
Qy 181 TDSKAAARLGSDMGGEFYMPYGTTRINQETVSLDANGVSGSASYEVKFSKPKNGQIWT 240
Db 188 TDSKAAARLGSDMGGEFYMPYGTTRINQETVSLDANGVSGSASYEVKFSKPKNGQIWT 247
Qy 241 GVIGSPAANADAGPPQRFVWVLGTANNPVVDKGAALAEISIRPLVAPPAPAPAPAP 300
Db 248 GVIGSPAANADAGPPQRFVWVLGTANNPVVDKGAALAEISIRPLVAPPAPAPAPAP 307
Qy 301 APAPAPAGEVAPTPTTPTPQRTLPA 325

Thu Jul 8 11:11:31 2004

us-10-720-192-3.rai

Page 8

Db 308 APAPAPAGEVAFTPTTPTQRTLPA 332
|||||

Search completed: July 7, 2004, 18:24:53
Job time : 17.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 18:24:06 ; Search time 42.5 Seconds
(without alignments)
2380.406 Million cell updates/sec

Title: US-10-720-192-3

Perfect score: 1726

Sequence: 1 MHQVDNLTERRKGRLLAALAI.....PAGEVAPPTPTPTQRTLPA 325

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1726	100.0	325	10	US-09-985-372-2
2	1726	100.0	325	10	US-09-985-372-3
3	1726	100.0	332	12	US-09-886-349A-41
4	1726	100.0	332	14	US-10-193-002-53
5	1726	100.0	332	14	US-10-084-843-53
6	1726	100.0	332	14	US-10-098-732A-41
7	1542	89.3	286	10	US-09-985-372-4
8	1499	86.8	652	14	US-10-193-002-350
9	1499	86.8	652	14	US-10-084-843-355
10	1499	86.8	802	9	US-09-287-849-10
11	1499	86.8	802	14	US-10-193-002-209
12	1499	86.8	802	14	US-10-193-002-346
13	1499	86.8	802	14	US-10-084-843-214
14	1499	86.8	802	14	US-10-084-843-351
15	1499	86.8	802	14	US-10-359-460-10

```

16 248.5 14.4 273 16 US-10-437-963-184760 Sequence 184760,
17 241 14.0 402 16 US-10-437-963-141313 Sequence 141313,
18 240.5 13.9 437 16 US-10-437-963-168762 Sequence 168762,
19 239 13.8 431 16 US-10-437-963-204963 Sequence 204963,
20 238 13.8 1269 16 US-10-437-963-130985 Sequence 130985,
21 237.5 13.8 697 12 US-10-425-114-41545 Sequence 41545, A
22 233.5 13.5 547 16 US-10-437-963-102579 Sequence 102579,
23 232.5 13.5 270 16 US-10-437-963-130068 Sequence 130068,
24 232.5 13.5 360 16 US-10-437-963-179473 Sequence 179473,
25 230.5 13.4 426 16 US-10-437-963-142229 Sequence 142229,
26 230.5 13.4 584 14 US-10-156-761-12405 Sequence 12405, A
27 229.5 13.3 569 16 US-10-437-963-149276 Sequence 149276,
28 229 13.3 440 16 US-10-437-963-138104 Sequence 128104,
29 228.5 13.2 235 16 US-10-437-963-162861 Sequence 162861,
30 227.5 13.2 652 15 US-10-104-047-3364 Sequence 3364, A
31 227 13.2 598 12 US-10-282-122A-84954 Sequence 64954, A
32 226.5 13.1 504 16 US-10-437-963-128105 Sequence 128105,
33 223.5 12.9 299 16 US-10-437-963-134054 Sequence 134054,
34 223.5 12.9 501 16 US-10-437-963-161137 Sequence 161137,
35 223.5 12.9 533 16 US-10-437-963-167716 Sequence 167716,
36 221.5 12.8 556 16 US-10-437-963-137918 Sequence 137918,
37 220.5 12.8 406 16 US-10-437-963-122256 Sequence 122256,
38 220 12.7 389 12 US-10-425-114-56400 Sequence 56400, A
39 220 12.7 466 16 US-10-437-963-195119 Sequence 195119,
40 219.5 12.7 369 16 US-10-437-963-196554 Sequence 196554,
41 219.5 12.7 376 16 US-10-437-963-154970 Sequence 154970,
42 218.5 12.7 285 16 US-10-437-963-189482 Sequence 189482,
43 218.5 12.7 534 16 US-10-437-963-159955 Sequence 159955,
44 217.5 12.6 350 16 US-10-437-963-102583 Sequence 102583,
45 217.5 12.6 817 15 US-10-369-493-1813 Sequence 1813, Ap

```

ALIGNMENTS

RESULT 1

```

US-09-985-372-2
; Sequence 2, Application US/0985372
; Publication No. US20030054008A1
; GENERAL INFORMATION:
; APPLICANT: LAQUEYERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICROORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 0660-0138-ODIV
; CURRENT APPLICATION NUMBER: US/09/985,372
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-2

```

```

Query Match 100.0%; Score 1726; DB 10; Length 325;
Best Local Similarity 100.0%; Pred. No. 8.6e-106;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MHQVDNLTERRKGRLLAALAIAMASASLVTVAVPATANADPEFPVPTTAASPPSTAAA 60
Db 1 MHQVDNLTERRKGRLLAALAIAMASASLVTVAVPATANADPEFPVPTTAASPPSTAAA 60
Qy 61 PPAPATVPAPPPAAAANTNAQPGDNNAPPPADPNAPPPVIAPNAPQVTRIDNPVGGF 120
Db 61 PPAPATVPAPPPAAAANTNAQPGDNNAPPPADPNAPPPVIAPNAPQVTRIDNPVGGF 120
Qy 121 SFALPAGWVESDAAHFDYGSALLSKTTGDPFPFGQPPPVANDTRIVGLRDLQKLYASABA 180

```



```

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guiderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 41
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DPEP
US-10-098-732A-41

Query Match      100.0%; Score 1726; DB 14; Length 332;
Best Local Similarity 100.0%; Pred. No. 8.7e-106;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHQVDPNLTTRKRLAALAIAAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 60
DB 8 MHQVDPNLTTRKRLAALAIAAMASASLVTVAVPATANADPEPAPPVPTTAASPPSTAAA 67

QY 61 PPAPATVAPPAPPAANTENAPGDPNAPADNAPPPVPIAPNAPQVRIQNPVGGF 120
DB 68 PPAPATVAPPAPPAANTENAPGDPNAPADNAPPPVPIAPNAPQVRIQNPVGGF 127

QY 121 SPALPAGWVESDAHFYDGSALLSKTTGDPFPFGPPQPPVANDTRIVLGRDQKLYASABA 180
DB 128 SPALPAGWVESDAHFYDGSALLSKTTGDPFPFGPPQPPVANDTRIVLGRDQKLYASABA 187

QY 181 TDSKAARLGSDMGDFYVPGTRINQETVSLDANGVSGSASYEVKFSKNGQIWT 240
DB 188 TDSKAARLGSDMGDFYVPGTRINQETVSLDANGVSGSASYEVKFSKNGQIWT 247

QY 241 VIGVSPAAANAPDAGPQRFVFWLGTANNPVDKGAALAESIRPLVAPPAPAPAPAE 300
DB 248 VIGVSPAAANAPDAGPQRFVFWLGTANNPVDKGAALAESIRPLVAPPAPAPAPAE 307

QY 301 APAPAPAGEVAPTPTTPTPQRTLPA 325
DB 308 APAPAPAGEVAPTPTTPTPQRTLPA 332

RESULT 7
US-09-985-372-4
; Sequence 4, Application US/09985372
; Publication No. US20030054008A1
; GENERAL INFORMATION:
; APPLICANT: LAQUEYRIERIE, Anne
; APPLICANT: MARCHAL, Gilles
; APPLICANT: PESCHER, Pascale
; APPLICANT: ROMAIN, Felix
; TITLE OF INVENTION: MYCOBACTERIAL PROTEINS, MICRORGANISMS PRODUCING THEM
; TITLE OF INVENTION: AND THEIR USE FOR VACCINES AND FOR THE DETECTION OF
; FILE REFERENCE: 0660-0138-0DIV
; CURRENT APPLICATION NUMBER: US/09/985,372
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/132,528
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 286
; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis
US-09-985-372-4

Query Match      89.3%; Score 1542; DB 10; Length 286;
Best Local Similarity 100.0%; Pred. No. 9.3e-94;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 DPEPAPPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTPNAPQGDPAAPADFNAPP 99
DB 1 DPEPAPPVPTTAASPPSTAAAPAPATPVAPPPPPAAANTPNAPQGDPAAPADFNAPP 60

QY 100 PVVIAENAPQVRIQNPVGGFSPALPAGWVESDAHFYDGSALLSKTTGDPFPFGQPPPV 159
DB 61 PVVIAENAPQVRIQNPVGGFSPALPAGWVESDAHFYDGSALLSKTTGDPFPFGQPPPV 120

QY 160 ANDTRIVLGRDQKLYASAEATDSKAAARLGSDMGDFYVPGTRINQETVSLDANGVSG 219
DB 121 ANDTRIVLGRDQKLYASAEATDSKAAARLGSDMGDFYVPGTRINQETVSLDANGVSG 180

QY 220 SASIYEVKFSKNGQIWTGIVGSPAANADAGPQRFVFWLGTANNPVDKGAAL 279
DB 181 SASIYEVKFSKNGQIWTGIVGSPAANADAGPQRFVFWLGTANNPVDKGAAL 240

QY 280 AESIRPLVAPPAPAPAPAEPAAPAPAPAGEVAPTPTTPTPQRTLPA 325
DB 241 AESIRPLVAPPAPAPAEPAAPAPAPAGEVAPTPTTPTPQRTLPA 286

RESULT 8
US-10-193-002-350
; Sequence 350, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 350:
; SEQUENCE CHARACTERISTICS:

```

```
;
; LENGTH: 652 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 350:
US-10-193-002-350

Query Match      86.8%; Score 1499; DB 14; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.5e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQCGDPAAPADPNAPPPVIAPN 106
DB 374 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQCGDPAAPADPNAPPPVIAPN 433
QY 107 APQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFQPPPVANDTRIV 166
DB 434 APQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFQPPPVANDTRIV 493
QY 167 LGRLDQKLYASAEATDSKAAARLGSMDGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 226
DB 494 LGRLDQKLYASAEATDSKAAARLGSMDGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 553
QY 227 KFSDPKPKNGQIWTGVIGSPAANAPDAGPPQRFVWMLGTANNPVDKGAALAESIRPL 286
DB 554 KFSDPKPKNGQIWTGVIGSPAANAPDAGPPQRFVWMLGTANNPVDKGAALAESIRPL 613
QY 287 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTPTPTORTLPA 325
DB 614 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTPTPTORTLPA 652

RESULT 9
US-10-084-843-355
; Sequence 355 Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Hedvick, Thomas S.
; Twardzik, Daniel R.
; Iodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
```

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 355:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 355:
US-10-084-843-355

Query Match      86.8%; Score 1499; DB 14; Length 652;
Best Local Similarity 100.0%; Pred. No. 1.5e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQCGDPAAPADPNAPPPVIAPN 106
DB 374 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQCGDPAAPADPNAPPPVIAPN 433
QY 107 APQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFQPPPVANDTRIV 166
DB 434 APQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFFQPPPVANDTRIV 493
QY 167 LGRLDQKLYASAEATDSKAAARLGSMDGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 226
DB 494 LGRLDQKLYASAEATDSKAAARLGSMDGEFFYMPYPGTRINQETVSLDANGVSGSASYEV 553
QY 227 KFSDPKPKNGQIWTGVIGSPAANAPDAGPPQRFVWMLGTANNPVDKGAALAESIRPL 286
DB 554 KFSDPKPKNGQIWTGVIGSPAANAPDAGPPQRFVWMLGTANNPVDKGAALAESIRPL 613
QY 287 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTPTPTORTLPA 325
DB 614 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTPTPTORTLPA 652

RESULT 10
US-09-287-849-10
; Sequence 10 Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-10
```



```
; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 346:
US-10-193-002-346

Query Match      86.8%; Score 1499; DB 14; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQGDPNAAAPPADPNAPPPVIAPN 106
DB 524 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQGDPNAAAPPADPNAPPPVIAPN 583
QY 107 APOQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 166
DB 584 APOQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 643
QY 167 LGRLDQKLYASAEATDSKAAARLGSMDGFEFYMPYPTGTRINQETVSLDANGVSGSASYEV 226
DB 644 LGRLDQKLYASAEATDSKAAARLGSMDGFEFYMPYPTGTRINQETVSLDANGVSGSASYEV 703
QY 227 KFSDSPKNGQIWTGVIGSPAANAPDAGPPQRFVFWVLGTANNPVDKGAALAESIRPL 286
DB 704 KFSDSPKNGQIWTGVIGSPAANAPDAGPPQRFVFWVLGTANNPVDKGAALAESIRPL 763
QY 287 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTTPQRTPLA 325
DB 764 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTTPQRTPLA 802

RESULT 13
US-10-084-843-214
; Sequence 214, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

; INFORMATION FOR SEQ ID NO: 346:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 802 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 346:
US-10-193-002-346

Query Match      86.8%; Score 1499; DB 14; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQGDPNAAAPPADPNAPPPVIAPN 106
DB 524 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQGDPNAAAPPADPNAPPPVIAPN 583
QY 107 APOQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 166
DB 584 APOQVRIDNPVGGFSFALPAGWVESDAAHFDYGSALLSKTTGDPFPQPPVANDTRIV 643
QY 167 LGRLDQKLYASAEATDSKAAARLGSMDGFEFYMPYPTGTRINQETVSLDANGVSGSASYEV 226
DB 644 LGRLDQKLYASAEATDSKAAARLGSMDGFEFYMPYPTGTRINQETVSLDANGVSGSASYEV 703
QY 227 KFSDSPKNGQIWTGVIGSPAANAPDAGPPQRFVFWVLGTANNPVDKGAALAESIRPL 286
DB 704 KFSDSPKNGQIWTGVIGSPAANAPDAGPPQRFVFWVLGTANNPVDKGAALAESIRPL 763
QY 287 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTTPQRTPLA 325
DB 764 VAPPPAPAPAEAPAPAPAPAGEVAPTPTTTPQRTPLA 802

RESULT 14
US-10-084-843-351
; Sequence 351, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
```

```
/ APPLICATION NUMBER: US/09/072,967
/ FILING DATE: 05-MAY-1998
/ ATTORNEY/AGENT INFORMATION:
/   NAME: Maki, David J.
/   REGISTRATION NUMBER: 31,392
/   REFERENCE/DOCKET NUMBER: 210121.411C9
/ TELECOMMUNICATION INFORMATION:
/   TELEPHONE: (206) 622-4900
/   TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 351:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 802 amino acids
/     TYPE: amino acid
/     STRANDEDNESS: single
/     TOPOLOGY: linear
/     MOLECULE TYPE: protein
/   SEQUENCE DESCRIPTION: SEQ ID NO: 351:
US-10-084-843-351

Query Match      86.8%; Score 1499; DB 14; Length 802;
Best Local Similarity 100.0%; Pred. No. 1.9e-90;
Matches: 279; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPN 106
Db 524 VPTTAASPPSTAAAPAPATPVAPPPAAANTPNAQPGDPNAAAPPADPNAPPPVIAPN 583
QY 107 APOQVRIDNPVGGFSFALPAGWVESDAAHFYGSALLSKTTGDPFPFGQPPPVANDTRIV 166
Db 584 APOQVRIDNPVGGFSFALPAGWVESDAAHFYGSALLSKTTGDPFPFGQPPPVANDTRIV 643
QY 167 LGRLDQKLYASAEATDSKAAARLGSDMGFEFYMPYPTGTRINQETVSLDANGVSGSASYEV 226
Db 644 LGRLDQKLYASAEATDSKAAARLGSDMGFEFYMPYPTGTRINQETVSLDANGVSGSASYEV 703
QY 227 KFSDSPKPNQGIWTGVIIGSPAANADAGPPQRFVVMWLTANNPVVDKGAAKALAESIRPL 286
Db 704 KFSDSPKPNQGIWTGVIIGSPAANADAGPPQRFVVMWLTANNPVVDKGAAKALAESIRPL 325
QY 287 VAPPPAPAPAPAPAPAPAPAGEVAPTPTTPTPQRTLPA 325
Db 764 VAPPPAPAPAPAPAPAPAPAGEVAPTPTTPTPQRTLPA 802

Search completed: July 7, 2004, 18:33:52
Job time : 61.5 secs
```

RESULT 15

```
US-10-359-460-10
/ Sequence 10, Application US/10359460
/ Publication NO. US20030147911A1
/ GENERAL INFORMATION:
/   APPLICANT: Reed, Steven G.
/   APPLICANT: Skeiky, Yasir A.W.
/   APPLICANT: Dillon, Davin C.
/   APPLICANT: Alderson, Mark
/   APPLICANT: Campos-Neto, Antonio
/   APPLICANT: Corixa Corporation
/   TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
/   TITLE OF INVENTION: and Their Uses
/   FILE REFERENCE: 014058-009020US
/   CURRENT APPLICATION NUMBER: US/10/359,460
/   CURRENT FILING DATE: 2003-02-05
/   PRIOR APPLICATION NUMBER: US/09/287,849
/   PRIOR FILING DATE: 1999-04-07
/   PRIOR APPLICATION NUMBER: US 08/818,112
/   PRIOR FILING DATE: 1997-03-13
/   PRIOR APPLICATION NUMBER: US 08/942,578
/   PRIOR FILING DATE: 1997-10-01
/   PRIOR APPLICATION NUMBER: US 09/025,197
/   PRIOR FILING DATE: 1998-02-18
/   PRIOR APPLICATION NUMBER: US 09/056,556
/   PRIOR FILING DATE: 1998-04-07
/   PRIOR APPLICATION NUMBER: US 09/223,040
/   PRIOR FILING DATE: 1998-12-30
/   NUMBER OF SEQ ID NOS: 46
```